On the Maximal Mediated Set Structure

and

the Applications of Nonnegative Circuit Polynomials

Von der

Carl-Friedrich-Gauß-Fakultät

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Oğuzhan Yürük

geboren am 23.06.1993

in Bursa, Turkei

Eingereicht am: 15.12.2020

Disputation am: 07.05.2021

1. Referentin/Referent: Prof. Dr. Timo de Wolff

2. Referentin/Referent: Prof. Dr. Anne Shiu

2021

Deutsche Zusammenfassung

Naiv gesehen ist ein Polynom nichts anderes als eine Reihe von Additionen und Multiplikationen. Man kann daher beim Arbeiten mit beliebigen abstrakten Strukturen, auf denen eine Addition und Multiplikation definiert sind, auf Polynome stoßen. Demnach tauchen Polynome in verschiedensten Bereichen der Mathematik auf und haben eine lange Geschichte in der Mathematik. Der Bereich der algebraischen Geometrie entstand insbesondere aus der Untersuchung von Lösungen von polynomiellen Gleichungssystemen, siehe [Die85] für einen umfassenden historischen Überblick. Polynome sind auch für Anwendungen der Mathematik in den Naturwissenschaften wichtig, da sie zur Darstellung von Beziehungen zwischen wissenschaftlich wichtigen Größen verwendet werden können. Zum Beispiel ist in der klassischen Physik die Flugbahn eines Projektils durch ein Polynom zweiten Grades gegeben, oder in der Chemie beschreibt die Massenwirkungskinetik die Geschwindigkeit einer Reaktion als Monom, d.h. als Polynom mit einem Term.

Die reelle algebraische Geometrie ist ein Bereich der Mathematik, der sich mit denjenigen Teilmengen von $\mathbb{R}^n$ befasst, die über Polynomgleichungen und -ungleichungen mit reellen Koeffizienten und Variablen definiert sind. Ein reelles Polynom wird als *nichtnegativ* über $\mathbb{R}^n$ bezeichnet, wenn dessen Auswertung an jedem beliebigen Punkt im $\mathbb{R}^n$ einen nichtnegativen Wert ergibt. Die Untersuchung der Nichtnegativität reeller multivariater Polynome ist nicht nur ein Schlüsselproblem in der reellen algebraischen Geometrie, sondern auch in der polynomiellen Optimierung ([Las10],[BPT12]) und in der Theorie chemischer Reaktionsnetzwerke ([CFMW17],[FKdWY20], [GH86], [EKW00], [HLS96]).

Die gebräuchlichste Methode, die Nichtnegativität eines Polynoms f zu zeigen, ist, es als *Summe von Quadraten* anderer Polynome zu schreiben, ist nicht a priori impliziert, dass f nicht negativ ist, siehe z.B. [Mar08], [Las10], [BPT12]. Im Jahr 1888 zeigte Hilbert [Hil88], dass die Darstellbarkeit als Summe von Quadraten keine notwendige Bedingung für Nichtnegativität ist – d.h. eines Polynoms dass nichtnegative Polynome existieren, die keine Summen von Quadraten sind – indem er ein nichtkonstruktives Gegenbeispiel angab. Später, in seiner berühmten Ansprache an den Internationalen Kongress der Mathematiker in Paris 1900, stellte er als sein 17. Problem eine Verallgemeinerung seiner früheren Ergebnisse, siehe [Hil00]. Diese Frage wurde von Emil Artin in [Art27] nach 27 Jahren positiv beantwortet, aber die Auswirkungen von [Hil88] und [Hil00] machten die

Nichtnegativität von reellen Polynomen und Summen von Quadraten zu einem aktiven Forschungsgebiet. Wir empfehlen [Rez00] für einen historischen Überblick zu Hilberts 17. Problem.

Das erste konkrete Beispiel eines nichtnegativen Polynoms, welches keine Summe von Quadraten ist, wurde 1967 in [Mot67] von Motzkin gegeben, siehe Example 2.3.1. Die Nichtnegativität dieses Polynoms bewies Motzkin mit Hilfe der *Ungleichung vom arithmetischen und geometrischen Mittel*, kurz *AM-GM-Ungleichung*, die besagt:

$$\sum_{j=0}^{d} \lambda_j t_j - \prod_{j=0}^{d} t_j^{\lambda_j} \geq 0,$$

für $t_j \geq 0$, $\lambda_j \geq 0$ und $\sum_{j=1}^{d} \lambda_j = 1$, siehe z.B. [Ste10, Kapitel 2]. Reznick beschäftigte sich eingehend mit den Formen, die sich aus der AM-GM-Ungleichung ergeben, und führte in [Rez89] die AGI-Formen – eine Klasse von Polynomen, deren Nichtnegativität aus der AM-GM-Ungleichung folgt – ein. In [IdW16a] verallgemeinerten Iliman und de Wolff die simplizialen AGI-Formen zu einer größeren Klasse von Polynomen, welche sie als *circuit polynomiale* bezeichnen. Außerdem bewiesen sie eine einfache Charakterisierung für die Nichtnegativität von circuit polynomialen, siehe Theorem 2.4.3. Das Schreiben eines Polynoms f als Summe von nichtnegativen circuit polynomialen (oder SONC-Polynom [1]) zertifiziert die Nichtnegativität von f. In den letzten Jahren gewannen die circuit polynomiale an Popularität und wurden zu einem aktiven Forschungsthema der polynomiellen Optimierung und der reellen algebraischen Geometrie. Wir verweisen beispielsweise auf [IdW16b], [SdW18] und [DHNdW20] für Anwendungen von SONCs auf die globale polynomielle Optimierung, auf [DIdW19], [DIdW17], [DKdW18] für den Fall der polynomiellen Optimierung unter Nebenbedingungen, und auf [FdW19], [DNT18], [W.20] für eine unvollständige Liste von Arbeiten zur Theorie der SONCs als Kegel in $\mathbb{R}^n$.

In dieser Dissertation untersuchen wir sowohl die Theorie als auch die Anwendung von Nichtnegativitätszertifikaten aus der Perspektive von circuit polynomialen. Diese Dissertation ist eine kollektive Arbeit der Forschung, die der Autor während seines Studiums als Doktorand zunächst an der TU Berlin danach an der TU Braunschweig durchgeführt hat. Einige Teile dieser Arbeit wurden bereits veröffentlicht oder sind Teil eines laufenden Projekts. Der Inhalt von Kapitel 3 ist in [HRdWY20] enthalten und ist eine gemeinsame Arbeit mit Olivia Röhrig und Timo de Wolff. Insbesondere die Berechnungen in Abschnitt 3.3 wurden von Olivia Röhrig im Rahmen ihrer Masterarbeit ([Roe20]) durchgeführt. Der Inhalt von Abschnitt 4.3 ist in [FKdWY20] enthalten und ist eine gemeinsame Arbeit mit Elisenda Feliu, Nidhi Kaihnsa und Timo de Wolff.

Der Inhalt dieser Arbeit gliedert sich in zwei wesentliche Teile.

[1] Abkürzung für "sum of nonnegative circuit polynomial" im Englischen

Grundlagen von Maximal Mediated Mengen

Im ersten Teil der Arbeit geben wir eine umfassende Diskussion von maximal mediated Mengen, welche in Zusammenhang mit den Newton-Polytopen von circuit polynomialen stehen. Die Newton-Polytope von circuit polynomialen sind ganzzahlige Simplexe mit Eckpunkten in $(2\mathbb{Z})^n$, und wir verwenden den Begriff simplicial basin für die Bezeichnung der Eckenpunktmenge solcher ganzzahliger Simplexe. Die *maximal gemittelte Menge* (siehe Definition 3.1.2 und Definition 3.1.9) eines simplicial basin S mit Ecken $\mathrm{Vert}\,(S)$ in $(2\mathbb{Z})^n$ ist die größte Teilmenge M von Gitterpunkten in $\mathbb{Z}^n \cap S$, die die folgenden beiden Eigenschaften erfüllt:

1. $\mathrm{Vert}\,(S) \subset M$, und
2. wenn $p \in M$, dann existieren $q_1, q_2 \in (2\mathbb{Z})^n \cap M$ mit $p = \dfrac{1}{2}(q_1 + q_2)$.

Es ist nicht a priori klar, ob es für jedes simplicial basin eine eindeutige maximal gemittelte Menge gibt. Reznick bewies in [Rez89, Theorem 2.2] die Existenz und die Eindeutigkeit der mit Newton-Polytopen von AGI-Formen assoziierten maximal mediated Mengen, aber seine Ideen lassen sich auch auf circuit polynomiale übertragen, wie in [IdW16a, Theorem 5.2] und [IdW16b, Corollary 3.2] untersucht wurde. Wir führen die Begriffe gemittelte und maximal gemittelte Menge für simplicial basin entsprechend dieser Beobachtung von de Wolff und Iliman rigoros ein und liefern einen neuen Beweis für die Existenz und Eindeutigkeit maximal gemittelter Mengen. Reznick gab auch einen Algorithmus (Algorithm 3.1.12) zur Berechnung maximal gemittelter Mengen in [Rez89] an. Wir betrachten in dieser Arbeit einen anderen Algorithmus (Algorithm 3.1.14), dessen Idee auf Timo de Wolff zurückgeht, und liefern einen vollständigen Beweis für dessen Korrektheit.

Wir interessieren uns für die maximal mediated Mengen, welche sich aufgrund der in [IdW16a, Satz 5.2] gegebenen Charakterisierung aus den Newton-Polytopen von circuit polynomialen ergeben. Diese Charakterisierung besagt, dass ein nichtnegatives circuit polynomial f genau dann eine Summe von Quadraten ist, wenn der Träger von f in der maximal mediated Menge des Trägers von f enthalten ist. Als historischer Kommentar sei hier angemerkt, dass dieses Ergebnis bereits früher für den Spezialfall der AGI-Formen von Reznick in [Rez89, Korollar 4.9] nachgewiesen wurde. Die von de Wolff und Iliman gegebene Charakterisierung kann weiter auf SONC-Polynome mit Simplex-Newton-Polytop ausgedehnt werden, wie wir in Theorem 3.1.26 beweisen. Daher hängt die Frage, ob ein SONC-Polynom mit Simplex-Newton-Polytop eine Summe von Quadraten ist, von der maximal mediated Menge ab, die mit ihm assoziiert ist. Um die maximal mediated Mengen systematisch zu untersuchen, führen wir den Begriff *h-ratio* (siehe Definition 3.2.1) eines simplicial basins ein. Das h-ratio eines simplicial basins Δ gibt die Dichte der maximal mediated Menge innerhalb von $\mathrm{conv}\,\Delta \cap \mathbb{Z}^n$ an. Wir verwenden

das h-ratio, um simplicial basin mit unterschiedlicher Struktur ihrer maximal mediated Mengen voneinander zu unterscheiden. Wir studieren jene Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}^n$, welche das h-ratio jedes simplicial basins in $2\mathbb{Z}^n$ erhält, unter dem Namen *maximal gemittelte erhaltende Funktionen des* $\mathbb{R}^n$, und geben eine vollständige Charakterisierung dieser Funktionen in Theorem 3.2.6. Diese Charakterisierung führt zu Korollar 3.2.7, welches uns erlaubt, diejenigen simplicial basin zu identifizieren, deren zugehörige Gitter (wie in (3.2.1) angegeben) bis auf Permutation die gleiche Hermite-Normalform haben.

In Zusammenarbeit mit Olivia Röhrig initiieren wir in `POLYMAKE` eine groß angelegte Berechnung, in der wir die maximal gemittelte Menge jedes simplicial basins mit fester Dimension n und maximalem Gesamtgrad $2d$ für die in Abschnitt 3.3.4 beschriebenen Fälle berechnen und die berechneten maximal mediated Mengen in einer Datenbank speichern. Diese ist unter

https://polymake.org/downloads/MMS/

verfügbar.

Wir analysieren zunächst den Fall $n = 2$ unter Verwendung unserer Datenbank und zeigen, dass Reznicks Vermutung Conjecture 3.1.19 für jedes beliebige simplicial basin mit maximalem Grad $2d \leq 150$ gilt. Wir erhalten aus unserer Datenbank außerdem zwei h-ratio-Verteilungen für jedes feste n und $2d$. Zunächst verfolgen wir die h-ratiose für jedes simplicial basin für feste n und $2d$, was die Verteilung des h-ratioses über alle simplicial basin ergibt. Zweitens verfolgen wir die h-ratiose für jede Äquivalenzklasse (definiert über die Identifizierung der simplicial basin, die dieselben Gitter haben, gemäß Korollar 3.2.7), was die Verteilung des h-ratioses über diejenigen Gitter ergibt, die sich aus den simplicial basin ergeben. Unter Verwendung dieser Datenbank von maximal mediated Mengen liefern wir Hinweise auf die Verteilungen des h-ratioses über simpliziale Mengen und Gitter. Insbesondere zeigen wir, dass diese beiden Verteilungen unterschiedlich sind.

Symbolic SONC Certificates and Multistationarity in CRNT

Der zweite Teil der Arbeit befasst sich mit der Anwendung von SONC-Polynomen auf die Theorie der chemischen Reaktionsnetzwerke (*CRNT*). Insbesondere untersuchen wir den Begriff der *Multistationarität* aus der CRNT, d.h. die Existenz mehrerer stationärer Zustände in einem chemischen Reaktionsnetzwerk. Multistationarität ist ein wichtiges Konzept in der CRNT, unter anderem aufgrund seiner Beziehung zur zellulären Entscheidungsfindung [LK99, OTL$^+$04, XF03]. Die Multistationarität eines bestimmten Reaktionsnetzwerks hängt oft von der Reaktionsratenkonstante (siehe Definition 4.1.1) ab. In der CRNT-Literatur gibt es verschiedene Methoden um zu entscheiden, ob sich bei einer gegebenen Wahl von Parameterwerten Multistationarität ergibt, siehe

z.B. [Fei95, Fel15, WF13, MDSC12, CFRS07, CHW08, DBMP14, EFJK12]. Es ist jedoch eine sehr schwierige Aufgabe, genau zu bestimmen, für welche Parameterwerte die Multistationarität eintritt.

Wir betrachten ein Modell des Phosphorylierungs- und Dephosphorylierungszyklus, welcher ein entscheidender chemischer Prozess im menschlichen Körper ist, siehe [Coh89]. Darüber hinaus ist dieses Modell ein Baustein der MAPK-Kaskade, bei der es sich um Signalwege handelt, die eine Vielzahl stimulierter zellulärer Aktivitäten regulieren, siehe [HF96, QNKS07, HR17]. Wir betrachten den Fall des 2-Stellen Phosphorylierungszyklus', der den Fall modelliert, dass ein Protein zwei mögliche Positionen für das Auftreten von Phosphorylierung und Dephosphorylierung hat. Wir bezeichnen die drei Phosphoformen eines gegebenen Proteins S mit keiner, einer, und zwei phosphorylierten Stellen jeweils mit S_0, S_1, und S_2 und wir nehmen an, dass die Phosporylierungs- und Dephosporylierungsereignisse durch das Kinaseenzym E bzw. das Phosphataseenzym F vermittelt werden. Daraus ergibt sich der folgende Mechanismus [WS08, CM14]:

$$S_0 + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} ES_0 \xrightarrow{\kappa_3} S_1 + E \underset{\kappa_8}{\overset{\kappa_7}{\rightleftharpoons}} ES_1 \xrightarrow{\kappa_9} S_2 + E$$
$$S_2 + F \underset{\kappa_{11}}{\overset{\kappa_{10}}{\rightleftharpoons}} FS_2 \xrightarrow{\kappa_{12}} S_1 + F \underset{\kappa_5}{\overset{\kappa_4}{\rightleftharpoons}} FS_1 \xrightarrow{\kappa_6} S_0 + F.$$

Unter der Annahme von Massenwirkungskinetik wird die zeitliche Entwicklung der Reaktantenkonzentrationen durch ein System autonomer ODEs in $\mathbb{R}^9_{\geq 0}$ modelliert, siehe Gleichung (4.2.3). Das System besteht aus Polynomgleichungen, deren Koeffizienten skalare Vielfache eines der 12 positiven Parameter $\kappa_1, \ldots, \kappa_{12}$ sind. Darüber hinaus ist die Dynamik auf stöchiometrische Kompatibilitätsklassen der Dimension sechs beschränkt, die durch die Gesamtmenge an Kinase, Phosphatase und Substrat gekennzeichnet sind, welche dann als Parameter in die Studie aufgenommen werden. Die weiteren Einzelheiten über das System sind in Abschnitt 4.2.1 zu finden.

Gegenwärtig ist bekannt, dass die Anzahl positiver stationärer Vorgänge innerhalb einer stöchiometrischen Kompatibilitätsklasse entweder eins oder drei beträgt, wenn alle positiven stationären Vorgänge nicht-degeneriert sind [WS08, MHK04] (siehe Abschnitt 4.1.2 für die Definition des nicht-degenerierten stationären Zustände). Es hat sich ferner gezeigt, dass es Parameterwahlen gibt, für die es zwei asymptotisch stabile stationäre Vorgänge und einen instabilen stationären Zustand gibt [HR15], siehe auch [TF20]. Einige neuere Fortschritte geben Aufschluss darüber, wie diese qualitativen Eigenschaften von der Wahl der Parameter abhängen. In [CM14] geben die Autoren zwei rationale Funktionen $a(\boldsymbol{\kappa})$ und $b(\boldsymbol{\kappa})$ der Parameter $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{12})$ (siehe (4.2.13) unten) mit den folgenden Eigenschaften an: Das System hat einen positiven stationären Zustand in jeder stöchiometrische Kompatibilitätsklasse, wenn $a(\boldsymbol{\kappa}) \geq 0$ und $b(\boldsymbol{\kappa}) \geq 0$, und es erlaubt mindestens zwei stationäre Vorgänge in einer beliebigen stöchiometrischen Kom-

patibilitätsklasse, wenn $a(\boldsymbol{\kappa}) \geq 0$, siehe Abschnitt 4.2.1. Darüber hinaus werden in [FW12, BDG20] Bedingungen für das Vorliegen von drei positiven stationäre Zustände mit den Parametern $\kappa_1, \ldots, \kappa_{12}$ angegeben. Um die Anzahl der stationären Vorgänge zu verstehen, verwenden wir Proposition 4.2.4, was ein Sonderfall von [CFMW17, Corollary 2] ist, und studieren das Vorzeichen des Polynoms $p_{\boldsymbol{\eta}}(x)$, welches in (4.2.12) angegeben ist. Dies führt zu einer vollständigen Charakterisierung der Multistationaritätsregion in Bezug auf kinetische Parameter für den 2-Stellen Phosphorylierungszyklus.

Wir liefern zwei hinreichende Bedingungen für Monostationarität: Erstens geben wir eine polynomielle Ungleichung in $\boldsymbol{\kappa}$ an, die mit Hilfe der Diskriminantentheorie hergeleitet wird, siehe Theorem 4.3.1, und den Bereich der Monostationarität im Fall $a(\boldsymbol{\kappa}) = 0$ vollständig charakterisiert. Für die zweite Ungleichung betrachten wir eine relevante SONC-Zerlegung, um eine hinreichende Bedingung für die Nichtnegativität von $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ zu finden, siehe Theorem 4.3.5. Obwohl diese Ungleichungen keine notwendigen Bedingungen für die Monostationarität sind, liefert die letztere Ungleichung vorläufige Informationen über die Form der Multistationaritätsregion (Korollar 4.3.10).

Wir gehen auf den Fall $a(\boldsymbol{\kappa}) \geq 0$ und $b(\boldsymbol{\kappa}) < 0$ ein, der in [CM14] [CFMW17] offen gelassen wurde, und wir zeigen, dass Multistationarität bei einer geeigneten Wahl von $\boldsymbol{\kappa}$ (siehe Proposition 4.3.11) auftreten kann. Darüber hinaus liefern wir eine parametrische Darstellung der Grenze zwischen den Regionen der Mono- und Multistationarität, in der Corollary 4.3.10 von Theorem 4.3.5 eine entscheidende Rolle spielt. In Theorem 4.3.18 schließen wir, dass die Region der Multistationarität in den Parametern $\kappa_1, \ldots, \kappa_{12}$ eine offene und zusammenhängende Menge ist, und dass die Region der Monostationarität in $\mathbb{R}^{12}_{>0}$ abgeschlossen und ebenfalls zusammenhängend ist.

Wir merken an, dass die in Abschnitt 4.3.2 verwendete Methode zur Zertifizierung der Nichtnegativität von $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ die erste konkrete Anwendung von SONC-Polynomen in der Literatur ist. In Abschnitt 4.4 gehen wir noch einen Schritt weiter und wenden diese Methode auf den Fall der Phosphorylierung an drei Stellen an. Dazu beweisen wir zunächst in Proposition 4.4.1 ein Analogon zu Proposition 4.2.4; dann beschreiben wir in Theorem 4.4.2 eine Teilmenge der monostationären Region, die durch drei Ungleichungen verallgemeinerter Polynome gegeben ist. Wir sehen weiter, dass die Teilmenge, die wir beschreiben, nicht leer ist, indem wir einen expliziten Punkt dieser Teilmenge berechnen. Wir schließen die Diskussion mit zwei Vermutungen ab: Conjecture 4.4.4 und Conjecture 4.4.5, welche Teil andauernder Forschung sind.

Acknowledgements

First and foremost, I wish to express my deep gratitude to my advisor Timo de Wolff because this thesis would not be possible without his continuous support and endless patience. Throughout my Ph.D. years, his guidance did not only support me in academic issues, but also helped me out with social issues, and broadened my view on life. In this regard, he was more than just academic advisor to me, therefore I would like to thank him separately for his contributions on my character.

I thank every single mathematician who contributed to my education, especially to Ali Sinan Sertöz and Robert Szöke. In particular, I am deeply grateful to Ali Nesin who introduced me and many other young aspiring mathematicians to the beauty of mathematics. Most importantly, I would like to give a big thanks to Bernd Sturmfels for being such a successful academic matchmaker, and specifically for introducing me to Timo de Wolff. I owe thanks to Michael Joswig for fostering me under his group in TU Berlin, after my research group moved to TU Braunschweig. Furthermore, I am grateful to have Elisenda Feliu, Nidhi Kaihnsa and Olivia Röhrig as my coauthors, who accompanied my academic research throughout my Ph.D. studies.

I also would like to thank all of my past and current group members from TU Berlin and TU Braunschweig for their friendship and support. In particular, I owe a great deal to my colleague Janin Heuer because of her immense support in my thesis. I would like to give my gratitude to Birgit Kommander for answering my never ending questions about the submission process, and Marta Panizzut for her valuable advice. Moreover, I appreciate all of the long discussions I had with Marek Kaluba about mathematics and mathematical software. I further would express my thankfulness to Lars Kastner and Benjamin Lorenz for answering all of my naive questions on how to use a computer properly.

I want to thank my beloved father and mother for letting me chase my dreams and teaching me to always stand up for my ideas. And at last, even though the words will not be enough to express my gratitude, I thank to my significant other Candan Güdücü. She is the sole reason how I was able to stay sane through the unpleasant experience of writing a thesis during a pandemic. Thank you for always believing in me.

Contents

Chapter 1

Introduction

From a naive perspective, a polynomial is just a series of additions and multiplications. So, one can encounter polynomials while working with any abstract ring structure that has an addition and a multiplication defined on it. Henceforth, polynomials show up in various areas of mathematics, and have a long history in mathematics. In particular, the field of algebraic geometry emerged from studying the systems of polynomial equations, see [Die85] for a comprehensive historical overview. Polynomials are also important for applications of mathematics in science, since they can be used to represent relations between scientifically significant quantities. For example in classical physics, the trajectory of a projectile is given by a degree two polynomial, or in chemistry, the mass action kinetics expresses the rate of a chemical reaction as monomial, i.e. a polynomial with one term, of the reactant quantities.

Real algebraic geometry is a mathematical subject that deals with the subsets of $\mathbb{R}^n$ that are defined by polynomial equations and inequalities with real coefficients and variables. A real polynomial is called *nonnegative* over $\mathbb{R}^n$, if its evaluation on any point of $\mathbb{R}^n$ yields a nonnegative value. Studying the nonnegativity of real multivariate polynomials is not only a key problem in real algebraic geometry, but also in polynomial optimization ([Las10, BPT12]), and in the theory of chemical reaction networks([CFMW17, FKdWY20, GH86, EKW00, HLS96])

The most common way to show the nonnegativity of a polynomial f, is to write it as *sum of squares (SOS)* of other polynomials, which a priori implies that f is nonnegative, see e.g., [Mar08, Las10, BPT12]. In 1888 [Hil88], Hilbert showed that being sum of squares is not a necessary condition for nonnegativity, i.e., there exists nonnegative polynomials that are not sums of squares, by giving a nonconstructive counterexample. Later in his famous 1900 address to the International Congress of Mathematicians in Paris, he posed a generalization of his earlier results as his 17th problem, see [Hil00]. His question was answered affirmatively in [Art27] by Emil Artin after 27 years, but the impact of [Hil88]

and [Hil00] transformed the nonnegativity of real polynomials and sums of squares an active area of research. We recommend [Rez00] for an historical review of Hilbert's 17th problem.

The first concrete example of a nonnegative polynomial that is not sum of squares was given by Motzkin in 1967 [Mot67], see Example 2.3.1. The nonnegativity of the Motzkin's polynomial follows from the classical *AM-GM inequality* which states that:

$$\sum_{j=0}^{d} \lambda_j t_j - \prod_{j=0}^{d} t_j^{\lambda_j} \ \geq \ 0,$$

for $t_j \geq 0$, $\lambda_j \geq 0$ and $\sum_{j=1}^{d} \lambda_j = 1$, see for example, [Ste10, Chapter 2]. Reznick vastly studied the forms arising from AM-GM inequality, and introduced the AGI-forms, which are a class of polynomials that are nonnegative due to AM-GM inequality, in [Rez89]. In [IdW16a] de Wolff and Iliman generalized the simplicial AGI-forms to a larger class of polynomials, which they define as *circuit polynomials* (see Definition 2.4.1). Furthermore, they pointed out an easy necessary condition for nonnegativity of circuit polynomials, see Theorem 2.4.3. Writing a polynomial f as a sum of nonnegative circuit (SONC) polynomials certifies the nonnegativity of f. In the last few years, the circuit polynomials gained popularity and became an active research topic in polynomial optimization and real algebraic geometry. For example, see [IdW16b, SdW18] and [DHNdW20] for applications of SONCs to the global polynomial optimization, see [DIdW19, DIdW17, DKdW18] for the case of constrained polynomial optimization, and see further [FdW19, DNT18, W.20] for a not exhaustive list of works on the theory of SONC polynomials as a cone in $\mathbb{R}^N$.

In this thesis, we study both the theory and the applications of the nonnegativity certificates from the perspective of circuit polynomials. This thesis is a collective work of the research that was done by the author during his studies as a doctoral student in TU Braunschweig and TU Berlin. Some parts of this thesis were already published, or are part of an ongoing project. The content of Chapter 3 is contained in [HRdWY20], and is a joint work with Olivia Röhrig and Timo de Wolff. Especially, the computations in Section 3.3 has been done by Olivia Röhrig as a part her Master's studies ([Roe20]). Some of the content in Chapter 4 is a joint work with is a joint work with Elisenda Feliu, Nidhi Kaihnsa and Timo de Wolff, and especially the content of Section 4.3 is contained in [FKdWY20].

The contents of the thesis are divided into two main parts.

1.1 Investigated Problems and Results

Foundations of Maximal Mediated Sets

In the first part of the thesis, we give a comprehensive discussion of the maximal mediated sets associated to the Newton polytopes of circuit polynomials. The Newton polytopes of circuit polynomials are integral simplices with vertices in $(2\mathbb{Z})^n$, and we use the term *simplicial basin* to denote vertex set of such integral simplices. The maximal mediated set (see Definition 3.1.2 and Definition 3.1.9) of a simplicial basin S with vertices $\operatorname{Vert}(S)$ in $(2\mathbb{Z})^n$ is the largest subset M of lattice points in $\mathbb{Z}^n \cap \operatorname{conv}(S)$ satisfying the following two properties:

1. $\operatorname{Vert}(S) \subset M$, and
2. if $\boldsymbol{p} \in M$, then there exist $\boldsymbol{q_1}, \boldsymbol{q_2} \in (2\mathbb{Z})^n \cap M$ with $p = \frac{1}{2}(\boldsymbol{q_1} + \boldsymbol{q_2})$.

A priori, the existence of an unique maximal mediated set for a given simplicial basin is not clear. Reznick proved the existence and the uniqueness of the maximal mediated sets associated to Newton polytopes of AGI-forms in [Rez89, Theorem 2.2], but his ideas extends to circuit polynomials as observed in [IdW16a, Theorem 5.2] and [IdW16b, Corllary 3.2]. We rigorously introduce the notions mediated and maximal mediated sets for simplicial basins following the observation of de Wolff and Iliman, and provide a new proof of [Rez89, Theorem 2.2] in the context of simplicial basins in Section 3.1.1. Reznick also pointed out an algorithm (Algorithm 3.1.12) to compute maximal mediated sets in [Rez89]. However, we consider another algorithm Algorithm 3.1.14, which was earlier pointed out by Timo de Wolff, and give a full proof of its correctness.

We are interested in the maximal mediated set arising from the Newton polytopes of circuit polynomials due to the characterization given in [IdW16a, Theorem 5.2], which states that a nonnegative circuit polynomial f is a sum of squares if and only if the support of f is contained in the maximal mediated set of the support of f. As a historical remark, we note that this result was earlier proven for the special case of simplicial AGI-forms by Reznick in [Rez89, Corollary 4.9]. The characterization given by Iliman and de Wolff can further be extended to SONC polynomials with simplex Newton polytope as we prove in Theorem 3.1.26. Therefore, the question of whether a SONC polynomial with simplex Newton polytope is a sum of squares depends on the maximal mediated set associated to it, see Corollary 3.1.27.

In order to study the maximal mediated sets systematically, in Section 3.2, we introduce the term *h-ratio* (see Definition 3.2.1) of a simplicial basin. The h-ratio of a simplicial basin Δ indicates the density of the MMS inside $\operatorname{conv}\Delta \cap \mathbb{Z}^n$, and we use h-ratio to distinguish the simplicial basins with different MMS structure from each other.

We study those maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ that preserve the h-ratio of every simplicial basin in $2\mathbb{Z}^n$ under the name *maximal mediated preserving set functions of* $\mathbb{R}^n$, and give a full characterization of them in Theorem 3.2.6. This characterization leads to Corollary 3.2.7, which allow us to identify the simplicial basins whose associated lattices (given as in (3.2.1)) share the same Hermite normal form up to permutation.

We initiate a large scale computation in `POLYMAKE` in collaboration with Olivia Röhrig, in which we compute the maximal mediated set of every simplicial basin of fixed dimension n and maximal total degree $2d$ for the cases described in Section 3.3.4, and store the computed maximal mediated sets in a database which is available at:

https://polymake.org/downloads/MMS/

We first analyze the case $n = 2$ and address to Conjecture 3.1.19, which states that the h-ratio of a 2-simplicial basin is necessarily 0 or 1. Using our database, we verify that Conjecture 3.1.19 holds for any simplicial basin with maximal degree $2d \leq 150$. To analyze our database further, we obtain two h-ratio distributions for each fixed n and $2d$. First, we keep track of the h-ratios for every simplicial basin for fixed n and $2d$, which yields the distribution of h-ratio over all simplicial basins. Second, we keep track of the h-ratios for each equivalence class (defined by identifying the simplicial basins that share the same lattices according to Corollary 3.2.7), which yields the distribution of h-ratio over lattices that arise from simplicial basins. In Section 3.3.5, using this database of maximal mediated sets, we study the distributions of h-ratio over simplicial sets and lattices. In particular, we show that these two distributions are different.

Symbolic SONC Certificates and Multistationarity in CRNT

The second part of the thesis addresses to the applications of SONC polynomials to the Chemical Reaction Networks Theory (CRNT). In particular, we study the notion of *multistationarity* from CRNT, which is the existence of multiple steady states in a chemical reaction network. Multistationarity is an important concept in CRNT because of, for example, its relation to cellular decision making and switch-like responses to graded input [LK99, OTL$^+$04, XF03]. The multistationarity of a given reaction network often depends on the reaction rate constants (see Definition 4.1.1). There are various methods in the CRNT literature to decide whether multistationarity arises for a given choice of parameter values, e.g., [Fei95, Fel15, WF13, MDSC12, CFRS07, CHW08, DBMP14, EFJK12]. However, it is a very difficult task to determine exactly for which parameter values multistationarity is enabled.

We consider a simple model of phosphorylation and dephosphorylation cycle which is a crucial chemical process in the human body [Coh89]. Furthermore, this model a building

block of the MAPK (mitogen activated protein kinase) cascade, which are signaling pathways that regulate a wide variety of stimulated cellular activities [HF96, QNKS07, HR17]. We consider the case of a 2-site phosphorylation cycle, which models the case a where protein has two possible sites for phosphorylation and dephosphorylation to occur. We denote the three phosphoforms of a given protein S with zero, one and two phosphorylated sites by S_0, S_1 and S_2, respectively. We denote the enzyme kinase, which mediates the phosphorylation of S, with E, and the enzyme phosphatase, which mediates the dephosphorylation of S, with F and dephosphorylation. This gives rise to the following mechanism [WS08, CM14]:

$$S_0 + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} ES_0 \xrightarrow{\kappa_3} S_1 + E \underset{\kappa_8}{\overset{\kappa_7}{\rightleftharpoons}} ES_1 \xrightarrow{\kappa_9} S_2 + E$$
$$S_2 + F \underset{\kappa_{11}}{\overset{\kappa_{10}}{\rightleftharpoons}} FS_2 \xrightarrow{\kappa_{12}} S_1 + F \underset{\kappa_5}{\overset{\kappa_4}{\rightleftharpoons}} FS_1 \xrightarrow{\kappa_6} S_0 + F.$$

Under the assumption of mass-action kinetics, the evolution of the reactant concentrations over time is modeled by a system of autonomous ODEs in $\mathbb{R}^9_{\geq 0}$, see equation (4.2.3). The system consists of polynomial equations, whose coefficients are scalar multiples of one of twelve positive parameters $\kappa_1, \ldots, \kappa_{12}$. Furthermore, the dynamics are constrained to stoichiometric compatibility classes of dimension six, characterized by the total amounts of kinase, phosphatase and substrate, which then enter the study as parameters. The further details about the system can be found in Section 4.2.1.

Before our results, the number of positive steady states within a linear invariant subspace is known to be either one or three, if all positive steady states are nondegenerate [WS08, MHK04] (see Section 4.1.2 for the definition of nondegenerate steady state). Moreover, it has been shown that there are parameter choices for which there exist two asymptotically stable steady states and one unstable steady state [HR15], see also [TF20]. Some recent progress has shed some light on how these qualitative properties depend on the choice of parameters. In [CM14] the authors give two rational functions $a(\boldsymbol{\kappa})$ and $b(\boldsymbol{\kappa})$ on the parameters $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{12})$ (see (4.2.13) below), with the following properties: the system has one positive steady state in each stoichiometric compatibility class if $a(\boldsymbol{\kappa}) \geq 0$ and $b(\boldsymbol{\kappa}) \geq 0$, and the system has multiple steady states in some stoichiometric compatibility class if $a(\boldsymbol{\kappa}) < 0$. Furthermore, in [FW12, BDG20] conditions for the existence of three positive steady states involving the parameters $\kappa_1, \ldots, \kappa_{12}$ and some of the total amounts are given, see also [CF12]. In order to understand the number of steady states, we use Proposition 4.2.4 (which is a special case of [CFMW17, Corollary 2]), and study the sign of the polynomial $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ given in (4.2.12). This leads to a complete characterization of the multistationarity region in terms of kinetic parameters for 2-site phosphorylation cycle.

We provide two sufficient conditions for monostationarity: First, we acquire a poly-

nomial inequality in $\boldsymbol{\kappa}$ using the theory of discriminants, see Theorem 4.3.1, which completely characterizes the region of monostationarity when $a(\boldsymbol{\kappa}) = 0$. For the second inequality, we consider a relevant SONC decomposition to find a sufficient condition for nonnegativity of $p_{\boldsymbol{\eta}}(\boldsymbol{x})$, see Theorem 4.3.5. Although these inequalities are not necessary for monostationarity, the Theorem 4.3.1 yields Corollary 4.3.10 that gives preliminary information about the shape of the multistationarity region.

We address to the crucial case $a(\boldsymbol{\kappa}) \geq 0$ and $b(\boldsymbol{\kappa}) < 0$, which was left open in [CM14] [CFMW17], and we show that multistationarity can occur for a suitable choice of $\boldsymbol{\kappa}$ in Proposition 4.3.11. Furthermore, we provide a parametric representation of the boundary between the the regions of mono- and multistationarity, in which the Corollary 4.3.10 of Theorem 4.3.5 plays a crucial role. In Theorem 4.3.18, we conclude that the region of multistationarity in the parameters $\kappa_1, \ldots, \kappa_{12}$ is an open and connected set, and the region of monostationarity is closed in $\mathbb{R}^{12}_{>0}$ and connected. We note that the method we use in Section 4.3.2 to certify the nonnegativity of $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ is the first concrete application of SONC polynomials in the literature.

In Section 4.4, we take one further step, and apply this method to the case of 3-site phosphorylation. In order to do so, first we prove an analog of Proposition 4.2.4 in Proposition 4.4.1, then in Theorem 4.4.2 we describe a subset of monostationarity region given by three inequalities of generalized polynomials. We further see that the subset we describe is not empty by computing an explicit point from this subset. We conclude the discussion with giving two conjectures Conjecture 4.4.4 and Conjecture 4.4.5, which are still a part of an ongoing research.

1.2 The Structure of the Thesis

Chapter 2 is the preliminary chapter where we provide the general notation, basic definitions and results that will be required in the rest of this thesis. To start with, in Section 2.1, we set the notation that we use throughout the thesis, and recall some fundamental definitions about polynomials. Next, in Section 2.2, we introduce the essential notions of the thesis, which are the nonnegative polynomials and the nonnegativity certificates. Furthermore, we stress the importance of these notions on mathematical applications, and point out two particular approaches to nonnegativity certification. We first discuss sums of squares certificates, and cover general results in Section 2.3. Then in Section 2.4, we introduce another nonnegativity certificate based on AM-GM inequality, which is the main approach we investigate in this thesis.

Chapter 3 focuses on the study of the maximal mediated sets, which is a notion that connects the two nonnegativity certificates from Section 2.3 and Section 2.4. We give a detailed introduction to the maximal mediated sets of simplicial basins in Section 3.1. In

this introduction, we first define the maximal mediated sets through the simplicial basins. Then, we explain the significance of maximal mediated sets, discuss two algorithms that compute maximal mediated sets, and give a refined summary of the known results from [Rez89] and [IdW16a] in Section 3.1.1. In Section 3.1.2, we discuss some additional observations by giving a generalization of a central theorem from Section 3.1.1. Next, in Section 3.2, to study maximal mediated sets in a structured manner, we introduce two essential definitions:h-ratio and MMS preserving functions. Then, we prove Theorem 3.2.6 that characterizes all MMS preserving maps, and prove a significant corollary of this theorem, i.e., Corollary 3.2.7. This key corollary yields an equivalence relation of maximal mediated sets, which is essential for discussion of the database generated in [HRdWY20]. In Section 3.3, we explain the generation of this database and make a statistical analysis of the database. For the generation of database, we first explain our approach to enumerate all simplicial basins in Section 3.3.1, then classify them according to Corollary 3.2.7 in Section 3.3.1. In Section 3.3.3, we stress on an algorithm to compute MMS and point out an implementation done by Olivia Röhrig in `POLYMAKE`. We explain in Section 3.3.4 the setup of the large scale computation that we did in order to create the database. Lastly, we give an analysis of the MMS database in Section 3.3.5.

Chapter 4 consists of a concrete application of the nonnegative circuit polynomial to the chemical reaction networks theory (CRNT). We start with a revision of the general knowledge about CRNT in Section 4.1. More specifically, in Section 4.1.1 we remind the reader about the key definitions of CRNT, recall the mass action kinetics, and see how to express a chemical reaction network as an ODE system. In Section 4.1.2, we discuss the steady states of the reaction, and revise an important result (Theorem 4.1.12) that enables us to give arguments on the multistationarity by studying the sign of a relevant polynomial. We emphasize on a concrete example of a chemical reaction network, called phosphorylation cycle, in Section 4.2. This is the main example in which we utilize nonnegative circuit polynomials to certify the preclusion of multistationarity. First, we see how Theorem 4.1.12 translates to our case study in Section 4.2.1, and furthermore, we cite the previously known results on the subject. In Section 4.2.2, we revise some additional algebraic tools which are required to understand the study of multistationarity done in [FKdWY20] for the case of 2-site. We study the parameter regions of multistationarity and monostationarity of 2-site phosphorylation cycle in Section 4.3, and cover the main results proven in [FKdWY20]. In Section 4.3.1 and Section 4.3.2, we provide two methods to describe a set of parameters from the monostationarity region, based on discriminant and SONC polynomials, respectively. In Section 4.3.3 we give a parametric description of boundary between mono and multistationarity regions. Lastly, in Section 4.3.4, we show that the regions of mono and multistationarity describe a connected region in parameter space for the 2-site phosphorylation cycle. The circuit approach which we employ can also

be extended to the higher number of sites, as we discuss in Section 4.4. In Section 4.4, we extend the circuit approach to describe a region in the parameter space that guarantees the monostationarity for the 3-site phosphorylation cycle, and we state two conjectures, Conjecture 4.4.4 and Conjecture 4.4.5, for the higher site cases. Conjecture 4.4.4 and Conjecture 4.4.5 are examined in an ongoing follow up project together with the authors of [FKdWY20], in which we study the monostationarity in the n-site phosphorylation cycle.

In Chapter 5, we give a general summary the thesis by revisiting the achieved results, together with an overview about the problems that are left open, and the ones that are still under investigation as a part of an ongoing work.

Chapter 2

Preliminaries

We use the notation $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$, respectively for real numbers, integers and natural numbers. We denote the closed interval between two given real numbers r, s such that $r > s$ by $[s, r]$, and the list of nonnegative integers $\{0, 1, \ldots, n\}$ by $[n]$. We write vectors with bold characters, and we refer to the i-th entry of the vector $\boldsymbol{\alpha}$ as α_i. Given two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we denote their *inner product* with $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{n} x_i y_i$. A point $\boldsymbol{\alpha} \in \mathbb{N}^n$ is called *even* if all of its entries are even. A subset of $\mathbb{N}^n$ is an *even set* if it consists of even points only. Given a finite set $L \subseteq \mathbb{N}^n$, we refer the cardinality of L as $\#L$, and the list formed by L with lexicographical ordering, or shortly lex-ordering, as $[L]$. We denote the convex hull of L by $\operatorname{conv}(L)$ and the vertices of $\operatorname{conv}(L)$ by $\operatorname{Vert}(L)$. We call the convex hull of an affine independent set of vectors in $\mathbb{R}^n$ of cardinality $k+1$ a *k-dimensional simplex*. We define the *scaled standard simplex* as

$$\operatorname{New}(f) = \Delta_{2d}^{n} := \left\{ \sum_{i=1}^{n} 2d \cdot \lambda_i \cdot \boldsymbol{e_i} \;\middle|\; \sum_{i=1}^{k} \lambda_i = 1 \text{ and } \lambda_i \geq 0 \right\} \subset \mathbb{R}^n, \tag{2.0.1}$$

where $2d \in \mathbb{N}$ and $\boldsymbol{e_i}$ denotes the i-th standard unit vector in $\mathbb{R}^n$. Given a k-dimensional simplex Δ, we denote the vertices of Δ by $\operatorname{Vert}(\Delta)$. We call a subset $N \subset \mathbb{R}^n$ *convex* if for any given two $\boldsymbol{x}, \boldsymbol{y} \in N$, it holds that $t\boldsymbol{x} + (1-t)\boldsymbol{y} \in N$ for all $t \in [0, 1]$. Given an n by m matrix $A \in \mathbb{R}^{n \times m}$, we denote its rank by $\operatorname{rank}(A)$, its transpose by A^T. If A is a square matrix, then we write its trace as $\operatorname{tr}(A)$ and its determinant as $\det(A)$. We denote the scalar multiplication of a vector $\boldsymbol{v} \in \mathbb{R}^n$(or a matrix $A \in \mathbb{R}^{n \times m}$) with a given scalar $\alpha \in \mathbb{R}$ by $\alpha \boldsymbol{v}$ (or by αA).

The rest of this chapter is divided into 4 sections. We start by giving a short introduction to polynomials in Section 2.1. Next, we discuss the notion of nonnegativity and its relation to polynomial optimization in Section 2.2. Section 2.3 consists of the information about the cone of SOS polynomials which will be mainly required in Chapter 3. Lastly,

we finish this chapter by giving a revision of circuit polynomials and the SONC cone, which constitutes the backbone of this thesis, in Section 2.4.

2.1 A Concise Introduction to Polynomials

In this section, we introduce the basic notations and definitions about polynomials, which can be found in any introductory texts in algebra such as [Lan]. Let K be a field. A polynomial $f(x_1, \dots, x_n)$ in variables $x_1, \dots, x_n$ over the field K is a finite sum of the form

$$f(x_1, \dots, x_n) = \sum_{\boldsymbol{\alpha} \in A \subset \mathbb{N}^n} f_{\boldsymbol{\alpha}} \left(\prod_{i=1}^{n} x_i^{\alpha_i} \right) \tag{2.1.1}$$

where $A \subset \mathbb{N}^n$ is a finite subset, and $f_{\boldsymbol{\alpha}} \in K$ for all $\boldsymbol{\alpha} \in A_f$. The *the support* of the polynomial f is set $A_f := \{\boldsymbol{\alpha} \in \mathbb{N}^n : \ f_{\boldsymbol{\alpha}} \neq 0\}$. *The Newton polytope* of a polynomial f is the convex hull of its support, and we denote it by $\mathrm{New}\,(f)$.

Remark 2.1.1. Note that in order for f to be a polynomial, the support A_f of f has to be a subset of $\mathbb{N}^n$. One can also obtain more general structures such as Laurent polynomials by allowing A_f to contain negative entries, or exponential sums by substituting x_i with e^{x_i} for all $i \in [n]$ and allowing real entries in A_f. However, throughout this thesis we assume that $A_f \subset \mathbb{N}^n$, unless stated otherwise. ⬡

Each summand in a polynomial f is called a *term*, and is indexed by an exponent $\boldsymbol{\alpha}$ from the support A_f. Each term of f consists of two parts: a coefficient and a monomial. $f_{\boldsymbol{\alpha}} \in K$ is called the *coefficient of f at the exponent $\boldsymbol{\alpha}$*. Of course, the choice of the ground field K directly effects the algebraic structure of the polynomial ring. Throughout the thesis we mostly focus on real polynomials, so $f_{\boldsymbol{\alpha}}$ will generally lie in $\mathbb{R}$ or $\mathbb{R}_{\geq 0}$. The expression $\prod_{i=1}^{n} x_i^{\alpha_i}$ is called the *monomial corresponding to the exponent $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$*. For simplicity of notation, we will use the multi index notation for monomials, i.e., we denote the monomial corresponding to the exponent $\boldsymbol{\alpha}$ by $\boldsymbol{x}^{\boldsymbol{\alpha}}$. We rewrite the polynomial f in (2.1.1) with multi index notation as

$$f(x_1, \dots, x_n) = \sum_{\boldsymbol{\alpha} \in A \subset \mathbb{N}^n} f_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}. \tag{2.1.2}$$

We denote the set of n-variate polynomials over a field K by $K[x_1, \dots, x_n]$, or by $K[\boldsymbol{x}]$. Given two polynomials $f, g \in K[\boldsymbol{x}]$ with supports $A_f, A_g \subseteq \mathbb{N}^n$, one can add f and g as

follows:

$$(f+g)(x_1,\dots,x_n) = \sum_{\boldsymbol{\alpha}\in A_f\cup A_g} (f_{\boldsymbol{\alpha}} + g_{\boldsymbol{\alpha}})\boldsymbol{x}^{\boldsymbol{\alpha}}. \tag{2.1.3}$$

The multiplication of the polynomials f and g is computed as follows:

$$(f\cdot g)(x_1,\dots,x_n) = \sum_{\boldsymbol{\gamma}\in A_{f\cdot g}} \Big(\sum_{\substack{\boldsymbol{\gamma}=\boldsymbol{\alpha}+\boldsymbol{\beta} \\ \boldsymbol{\alpha}\in A_f \\ \boldsymbol{\beta}\in A_g}} f_{\boldsymbol{\alpha}}\cdot g_{\boldsymbol{\beta}} \Big) \boldsymbol{x}^{\boldsymbol{\gamma}}. \tag{2.1.4}$$

where

$$A_{f\cdot g} := \{\boldsymbol{\alpha}+\boldsymbol{\beta}\in\mathbb{N}^n : \ \boldsymbol{\alpha}\in A_f \text{ and } \boldsymbol{\beta}\in A_g\}.$$

Note that the addition and multiplication of the coefficients corresponds to the binary operations of the field K. It is a straightforward exercise to show that when K is a field $K[\boldsymbol{x}]$ has a commutative ring structure with the addition and multiplication defined as in (2.1.3) and (2.1.4).

The *total degree of the monomial* $\boldsymbol{x}^{\boldsymbol{\alpha}}$ is the 1-norm of the vector $\boldsymbol{\alpha}$, i.e. $||\boldsymbol{\alpha}||_1 := \sum_{i=1}^n \alpha_i$. *The total degree*, or simply *the degree of a polynomial* $f\in\mathbb{R}[\boldsymbol{x}]$ is then defined as the maximum total degree of its nonzero terms, and we denote it by $\deg(f)$. If $f\in\mathbb{R}[\boldsymbol{x}]$ consists of terms that have the same total degree, then f is called a *homogeneous polynomial*, or *an n-ary form*. By an elementary combinatorial argument, one can calculate for any $n,d\in\mathbb{N}$ that the number of monomials with n variables and total degree d equals to $\binom{n+d-1}{d}$. Then, we can calculate the number of monomials of degree at most d by simply taking a sum over the degree:

$$\sum_{i=0}^{d} \binom{n-1+i}{i} = \binom{n+d}{d}. \tag{2.1.5}$$

Alternatively, one can avoid computing this combinatorial sum by counting the set of $(n+1)$-variate monomials of degree exactly equal to d, which has the same cardinality as the set of n-variate monomials of degree at most d. Indeed, given a monomial $\prod_{i=1}^{n+1} x_i^{\alpha_i}$ such that $\sum_{i=1}^{n+1}\alpha_i = d$, one can find a unique monomial of degree at most d in variables $x_1,\dots x_n$ by setting $x_{n+1}=1$. Conversely, any n-variate monomial $\prod_{i=1}^{n} x_i^{\alpha_i}$ with $d' := \sum_{i=1}^n \alpha_i \le d$ can be uniquely extended to a $(n+1)$-variate monomial of degree exactly d as $\prod_{i=1}^n x_i^{\alpha_i} x_{n+1}^{d'-d}$. Hence, one can directly compute the right hand side of (2.1.5) by counting the number of $(n+1)$-variate monomials of degree exactly d, which is equal to $\binom{n+d}{d}$.

Sometimes, instead of working with entire the $\mathbb{R}[\boldsymbol{x}]$, we work with certain subsets of

$\mathbb{R}[\boldsymbol{x}]$. Given $n, d \in \mathbb{N}$, we consider the following basic subsets of $\mathbb{R}[\boldsymbol{x}]$:

- the set of polynomials in $\mathbb{R}[\boldsymbol{x}]$ of degree at most d, which we denote by $\mathbb{R}[\boldsymbol{x}]_d$,
- the set of real homogeneous polynomials in $\mathbb{R}[\boldsymbol{x}]$ (*forms*) of degree d, which we denote by $H_{n,d}$.

First, we point out that both $\mathbb{R}[\boldsymbol{x}]_d$ and $H_{n,d}$ are closed under the addition defined in (2.1.3), as well as multiplication by a scalar in $\mathbb{R}$. Thus, by considering each monomial as a basis vector, one can see that $\mathbb{R}[\boldsymbol{x}]_d$ and $H_{n,d}$ are real vector spaces of dimensions $\binom{n+d}{d}$ and $\binom{n+d-1}{d}$ over $\mathbb{R}$. For some purposes, it is more useful to start with a fixed set of exponents $A \subset \mathbb{N}^n$. Then, we consider the vector space generated by only those monomials $\boldsymbol{x}^{\boldsymbol{\alpha}}$ such that $\boldsymbol{\alpha} \in A$. Given $A \subset \mathbb{N}^n$, we define the *vector space of polynomials that are supported on* A as

$$\mathbb{R}^A := \{f \in \mathbb{R}[\boldsymbol{x}] : \ A_f \subseteq A\}. \tag{2.1.6}$$

Given a polynomial $f \in \mathbb{R}^A$, *the 1-norm of the polynomial* f is defined as

$$||f||_1 := \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}} \tag{2.1.7}$$

While we mostly use the language of polynomials, our results are transferable to the language of homogeneous forms. Any polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]_d$ can be homogenized into an form in $H(\mathbb{R}^{n+1})'_d$ with $d' \geq d$. In order to do so, one can introduce a new variable x_{n+1}, and define the following form:

$$x_{n+1}^{d'} \cdot f\left(\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}\right). \tag{2.1.8}$$

Note that one can define a new homogenization for each $d' \in \mathbb{N}$ that is greater than or equal to d. Unless it is stated otherwise, we always consider the homogenization with $d' = d$. Observe that we can recover the original polynomial f by setting $x_{n+1} = 1$ in (2.1.8). In general, any given form $h \in H_{n,d}$ can be dehomogenized into a polynomial in $\mathbb{R}[x_1, \ldots, x_{n-1}]_d$ in a similar manner, e.g. via dehomogenizing over the last entry by setting $x_n = 1$:

$$h(x_1, \ldots, x_{n-1}, 1). \tag{2.1.9}$$

We use homogenization and dehomogenization to carry over ideas, arguments and constructions between $\mathbb{R}[x_1, \ldots, x_n]_d$ and $H_{n+1,d}$. For example, the Newton polytope of any given $f \in \mathbb{R}[x_1, \ldots, x_n]_d$ is an embedded copy of the Newton polytope of its homogeniza-

tion in $H(\mathbb{R}^{n+1})_d$. Similarly, dehomogenizing a form $h \in H_{n,d}$ corresponds to projecting the Newton polytope $\mathrm{New}\,(h) \subset \mathbb{R}^n$ to an $(n-1)$-dimensional hypersurface via the map

$$(\alpha_1, \ldots, \alpha_n) \longmapsto (\alpha_1, \ldots, \alpha_{n-1}, 0)\,.$$

In the rest of this chapter we discuss some significant characteristics of polynomials and forms, which constitute the backbone of this thesis. Most of these characteristics are invariant under the homogenization and dehomogenization, which allows us to work with the sets $\mathbb{R}[\boldsymbol{x}]_d$ and $H_{n+1,d}$ interchangeably. We note here the discussion is given from the perspective of $\mathbb{R}[\boldsymbol{x}]_d$ throughout the thesis.

2.2 Nonnegative Polynomials and Polynomial Optimization

In this section, we discuss the nonnegative polynomials, and their relation to polynomial optimization. We refer to [Rez00] for a comprehensive historical overview about the nonnegative polynomials, and to [PD13, Mar08, Lau09, BPT12] for a detailed discussion.

A polynomial in $f \in \mathbb{R}[x_1, \ldots, x_n]$ or a form $f \in H_{n,d}$, contains more information than an arbitrary ring element. In particular, one can consider f as map from $\mathbb{R}^n$ to $\mathbb{R}$ by evaluating f at a given point $\boldsymbol{x} \in \mathbb{R}^n$. As a result of $\mathbb{R}$ being a totally ordered set, one can compare the elements in the image of f, i.e., $\mathrm{im}(f) = \{f(\boldsymbol{x}) \in \mathbb{R} \mid \boldsymbol{x} \in \mathbb{R}^n)\}$. This has various implications. In particular, it allows us to make the following definition:

Definition 2.2.1. A polynomial $f \in \mathbb{R}[\boldsymbol{x}]_d$ (or a form in $H_{n,d}$) is called *nonnegative*, if $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$. We denote the set of nonnegative polynomials in $\mathbb{R}[\boldsymbol{x}]_d$ and $H_{n,d}$ by $P\left(\mathbb{R}[\boldsymbol{x}]_d\right)$ and $P\left(H_{n,d}\right)$, respectively.

Similarly, an element f of $\mathbb{R}^A$ for some given $A \subset \mathbb{N}^n$, is called *nonnegative* if $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$, and we denote the set of nonnegative polynomials in $\mathbb{R}^A$ by $P\left(\mathbb{R}^A\right)$. ⬡

We first point out that if a polynomial $f \in P\left(\mathbb{R}[\boldsymbol{x}]_d\right)$, then the vertices of $\mathrm{New}\,(f)$ are necessarily even points in $\mathbb{N}^n$. This fact is a folklore in the literature, see, for example, the lemma in page 365 of [Rez78]. Therefore, we continue the discussion of nonnegativity with focusing on polynomials of even degree from now on. If $f \in P\left(\mathbb{R}[\boldsymbol{x}]_{2d}\right)$, then its homogenization to degree $2d' \geq 2d$, i.e.,

$$x_{n+1}^{2d'} \cdot f\left(\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}\right),$$

is also nonnegative. Moreover, if $h \in H_{n,d}$ is a nonnegative form, then dehomogenizing h

at any variable (as described for the last variable in equation (2.1.9)) yields an $(n-1)$-variate nonnegative polynomial.

$P(\mathbb{R}[\boldsymbol{x}]_{2d})$ and $P(\mathbb{R}^A)$ are very well structured subsets of $\mathbb{R}[\boldsymbol{x}]_{2d}$ and $\mathbb{R}^A$, respectively. In particular, given two $f, g \in P(\mathbb{R}[\boldsymbol{x}]_{2d})$ and $\alpha \in \mathbb{R}_{\geq 0}$, then it follows that $\alpha f + g \in P(\mathbb{R}[\boldsymbol{x}]_{2d})$, i.e., $P(\mathbb{R}[\boldsymbol{x}]_{2d})$ forms a cone in the vector space $\mathbb{R}[\boldsymbol{x}]_{2d}$. Similarly, one can see that the set $P(\mathbb{R}^A)$ also bears a cone structure in $\mathbb{R}^A$. In consequence, we call $P(\mathbb{R}[\boldsymbol{x}]_{2d})$ *the cone of positive polynomials*, and $P(\mathbb{R}^A)$ *the cone of positive polynomials supported on A*. Note that both of these cones are convex, because $tf + (1-t)g$ is in $P(\mathbb{R}[\boldsymbol{x}]_{2d})$ (or in $P(\mathbb{R}^A)$ for a given A) for all $f, g \in P(\mathbb{R}[\boldsymbol{x}]_{2d})$ (or in $P(\mathbb{R}^A)$ respectively) and $t \in [0,1]$. Next, in Proposition 2.2.2 we give an already established proof of a key fact about the cone $P(\mathbb{R}[\boldsymbol{x}]_{2d})$.

Proposition 2.2.2. $P(\mathbb{R}[\boldsymbol{x}]_{2d})$ is a closed convex cone in $\mathbb{R}[\boldsymbol{x}]_{2d} \simeq \mathbb{R}^N$ where $N = \binom{n+2d}{n}$.

Proof. We have already pointed out that $P(\mathbb{R}[\boldsymbol{x}]_{2d})$ is a convex cone, we now show that $P(\mathbb{R}[\boldsymbol{x}]_{2d})$ is closed. We proceed by showing that $\mathbb{R}[\boldsymbol{x}]_{2d} \setminus P(\mathbb{R}[\boldsymbol{x}]_{2d})$ is open. So, let $f \in \mathbb{R}[\boldsymbol{x}]_{2d} \setminus P(\mathbb{R}[\boldsymbol{x}]_{2d})$, and hence there exists $\boldsymbol{y_0} \in \mathbb{R}^n$ such that $f(\boldsymbol{y_0}) < 0$. For $\varepsilon \in \mathbb{R}_{>0}$, we define *the N-dimensional open ε-ball around f*,

$$B_\varepsilon(f) := \left\{ g = \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}^n \\ ||\boldsymbol{\alpha}||_1 \leq 2d}} g_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} \in \mathbb{R}[\boldsymbol{x}]_{2d} : \; g \in P(\mathbb{R}[\boldsymbol{x}]_{2d}) \text{ and } \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}^n \\ ||\boldsymbol{\alpha}||_1 \leq 2d}} ||g_{\boldsymbol{\alpha}} - f_{\boldsymbol{\alpha}}||_1 \leq \varepsilon \right\},$$

where $||g_{\boldsymbol{\alpha}} - f_{\boldsymbol{\alpha}}||_1$ denotes the one norm given as in (2.1.7).

Our aim is to find a suitable $\varepsilon > 0$ such that $B_\varepsilon(f) \subset \mathbb{R}[\boldsymbol{x}]_{2d} \setminus P(\mathbb{R}[\boldsymbol{x}]_{2d})$. Let

$$\mu := \max_{\substack{\boldsymbol{\alpha}\in\mathbb{N}^n \\ ||\boldsymbol{\alpha}||_1 \leq 2d}} ||\boldsymbol{y_0}^{\boldsymbol{\alpha}}||,$$

and fix an $\varepsilon < \frac{-f(\boldsymbol{y_0})}{2\mu}$. Then for any $g \in B_\varepsilon(f)$

$$\begin{aligned} g(\boldsymbol{y_0}) &= \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}^n \\ ||\boldsymbol{\alpha}||_1 \leq 2d}} ((g_{\boldsymbol{\alpha}} - f_{\boldsymbol{\alpha}}) + f_{\boldsymbol{\alpha}})\, \boldsymbol{y_0}^{\boldsymbol{\alpha}} \\ &= \sum_{\substack{\boldsymbol{\alpha}\in\mathbb{N}^n \\ ||\boldsymbol{\alpha}||_1 \leq 2d}} (g_{\boldsymbol{\alpha}} - f_{\boldsymbol{\alpha}})\boldsymbol{y_0}^{\boldsymbol{\alpha}} + f(\boldsymbol{y_0}) \leq \mu\varepsilon + f(\boldsymbol{y_0}) < \frac{f(\boldsymbol{y_0})}{2} < 0 \end{aligned}$$

Therefore, $g \in \mathbb{R}[\boldsymbol{x}]_{2d} \setminus P(\mathbb{R}[\boldsymbol{x}]_{2d})$, and consequently $B_\varepsilon(f) \subset \mathbb{R}[\boldsymbol{x}]_{2d} \setminus P(\mathbb{R}[\boldsymbol{x}]_{2d})$. This proves that $\mathbb{R}[\boldsymbol{x}]_{2d} \setminus P(\mathbb{R}[\boldsymbol{x}]_{2d})$ is open, or equivalently $\mathbb{R}[\boldsymbol{x}]_{2d}$ is closed. □

Note that the open ball $B_\varepsilon(f)$ that we constructed in the proof of Proposition 2.2.2 is full dimensional. Following a similar approach, given an $f \in P\left(\mathbb{R}[\boldsymbol{x}]_{2d}\right)$ and a $\delta \in \mathbb{R}_{>0}$, one can also show that there exists a full dimensional ball in $P\left(\mathbb{R}[\boldsymbol{x}]_{2d}\right)$ around $f+\delta$. We can further show, with the same argumentation, that $P\left(\mathbb{R}^A\right)$ is a closed convex cone for any given nonempty $A \subset \mathbb{N}^n$. We investigate the certain subcones of these nonnegativity cones in Section 2.3 and Section 2.4, and are particularly interested in deciding whether a given polynomial lie in one of these cones.

Given a set of polynomials $\mathscr{G} = \{g_1, \ldots, g_s\} \in \mathbb{R}[\boldsymbol{x}]$, we call the set

$$\mathscr{G}_+ := \{\boldsymbol{x} \in \mathbb{R}^n \mid g_k(\boldsymbol{x}) \geq 0 \text{ for all } g_k \in \mathscr{G}\}$$

a *basic closed semi-algebraic set*. We say that a polynomial $f \in \mathbb{R}[\boldsymbol{x}]$ is *nonnegative over* $\mathscr{G}_+$, if $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathscr{G}_+$.

Example 2.2.3. Two of the most elementary example of a basic closed semi-algebraic set are $\mathbb{R}^n$ and $\mathbb{R}^n_{\geq 0}$ for any positive integer n. For example, if we let $\mathscr{G}$ consist of a single constant polynomial $g(\boldsymbol{x}) = 1$, then $\mathscr{G}_+ = \mathbb{R}^n$. Similarly, if we consider the set of linear constraint functions $\mathscr{G} = \{g_1(\boldsymbol{x}), \ldots, g_n(\boldsymbol{x})\}$ such that $g_i(\boldsymbol{x}) = x_i$, then $\mathscr{G}_+ = \mathbb{R}^n_{\geq 0}$. ⬡

We are interested in checking whether f is nonnegative over a given semi-algebraic set $\mathscr{G}_+$. In particular, the notion of nonnegativity is closely related to polynomial optimization.

Definition 2.2.4. For $f, g_1, \ldots, g_s \in \mathbb{R}[x_1, \ldots, x_n]$, a *constrained polynomial optimization problem (CPOP)* is given as

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && g_1(\boldsymbol{x}), \ldots, g_s(\boldsymbol{x}) \geq 0 \end{aligned} \tag{2.2.1}$$

We call f the *objective function*, and the set of functions $\mathscr{G} = \{g_1, \ldots, g_s\}$ as the *constraint functions*. The region that is described by the constraint functions, i.e. $\mathscr{G}_+$, is called the *feasible region* of the CPOP given in (2.2.1). ⬡

Solving CPOPs is extremely useful for an immense number of applications in science, engineering and mathematics. There are already well established methods for convex optimization problems including but not limited to linear programming, geometric programming or semi-definite programming, see e.g. [BV11]. Unfortunately, one cannot directly use convex optimization methods to solve CPOPs. We recall here that a function f from $\mathbb{R}^n$ to $\mathbb{R}$ is called *convex function* if for any given two $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ and for all $t \in [0, 1]$

$$f(t\boldsymbol{v} + (1-t)\boldsymbol{w}) \leq tf(\boldsymbol{v}) + (1-t)f(\boldsymbol{w}), \tag{2.2.2}$$

and a convex optimization problem consists of convex objective functions and convex constraint functions. If we consider the univariate polynomial $f(x) = x^3+x^2$ (see Figure 2.1), then we see that $f(x)$ is not convex since (2.2.2) does not hold for $v = -2, w = 0, t = \frac{1}{4}$:

$$f(tv + (1-t)w) = -\frac{1}{8} > -1 = tf(v) + (1-t)f(w).$$

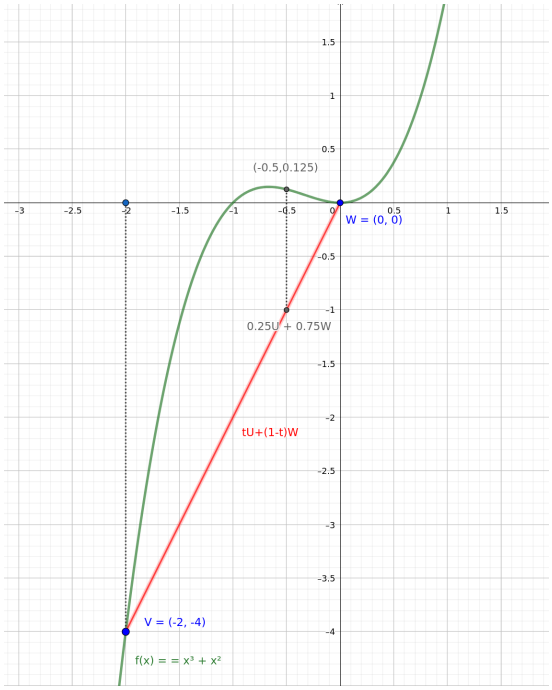

Figure 2.1: Green curve shows the graph of $f(x) = x^3 + x^2$, which is not a convex function. Two blue points denote $\boldsymbol{V} = (v, f(v)) = (0, 0)$ and $\boldsymbol{W} = (w, f(w)) = (-2, -4)$, respectively. The left hand side of (2.2.2) yields the red line segment between $\boldsymbol{V}$ and $\boldsymbol{W}$, which is below the graph of $f(x)$. Hence, the inequality in (2.2.2) fails.

As a consequence of CPOPs not being convex, solving CPOPs turns out to be a hard problem. In fact, the famous MAX-CUT problem can be formulated as a CPOP.

Example 2.2.5 (MAX-CUT)**.** Let $G := (V, E)$ be a finite graph, where vertices are labeled as $V := [k]$ for some $k \in \mathbb{N}$, and each edge $(i, j) \in E$ is given a weight $w_{ij} \in \mathbb{R}$. A *cut* in G is a set of edges induced by some subset S of vertices such that:

$$\{(i, j) \in E \mid i \in S \text{ and } j \in V \setminus S\}.$$

The weight of a given cut S is defined as $\sum_{(i,j)\in S} w_{ij}$. The MAX-CUT problem is the task

of finding the cut with the maximal weight, and it can be expressed as a CPOP. In order to do so, we first consider the $|V|$-dimensional *Boolean hypercube*, $\{\pm 1\}^k \subset \mathbb{R}^k$. Given a subset S of V, we set $x_i = 1$ if $i \in S$, and $x_i = -1$ otherwise. Then, the MAX-CUT problem can be reformulated as

$$\begin{aligned} \text{maximize} \quad & \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j) \cdot w_{ij} \\ \text{subject to} \quad & x_i^2 - 1 = 0 \text{ for all } i \in [k]. \end{aligned}$$

We also note that MAX-CUT is one of the problems that is in the Karp's famous list of NP-complete problems [Kar72]. The optimization version we pointed out here is NP-hard, see [GW95] for a polynomial-time approximation algorithm via semi definite programming. ⬡

We use the notion of nonnegativity to tackle this problem by reformulating CPOPs in terms of nonnegativity. Essentially, if $\gamma \in \mathbb{R}$ real number such that $f(\boldsymbol{x}) - \gamma \geq 0$ for all $\boldsymbol{x} \in \mathscr{G}_+$, then $f(\boldsymbol{x}) \geq \gamma$ for $\boldsymbol{x} \in \mathscr{G}_+$. Moreover, this means that

$$\min\{f(\boldsymbol{x}) \in \mathbb{R} \mid \boldsymbol{x} \in \mathscr{G}_+\} = \max\{\gamma \in \mathbb{R} \mid f(\boldsymbol{x} - \gamma \geq 0 \ \forall \boldsymbol{x} \in \mathscr{G}_+)\}. \tag{2.2.3}$$

Therefore, in order to solve a CPOP as given in Definition 2.2.4, it is enough to compute the maximum $\gamma \in \mathbb{R}$ such that $f(\boldsymbol{x}) - \gamma \geq 0$ for all $\boldsymbol{x} \in \mathscr{G}_+$. If the CPOP is given as a maximization problem, then, similarly, it is enough to compute the minimum $\gamma \in \mathbb{R}^n$ such that $-f(\boldsymbol{x}) + \gamma \geq 0$ for all $\boldsymbol{x} \in \mathscr{G}_+$. See Figure 2.2 for an illustration of this correspondence for the global optimization case , i.e. for $\mathscr{G}_+ = \mathbb{R}^n$.

Remark 2.2.6. For the sake of completeness, we have introduced polynomial optimization in the constrained setting. A CPOP is called an *global polynomial optimization problem* if its feasible region is $\mathbb{R}^n$. Throughout the thesis we mostly focus on global polynomial optimization, and therefore we are particularly interested in the cases where $\mathscr{G}_+ = \mathbb{R}^n$, and sometimes $\mathscr{G}_+ = \mathbb{R}^n_{\geq 0}$. ⬡

Even though the task of deciding the nonnegativity of an arbitrary polynomial $f \in \mathbb{R}[\boldsymbol{x}]$ over $\mathbb{R}^n$ or over a semi-algebraic subset of $\mathbb{R}^n$ is useful, it is a task hard both in theory and practice. In [Par00], Parrilo points out, by citing [MK87, Theorem 2], that certifying global positivity of a polynomial is NP-hard when the degree of the polynomial is at least four. Therefore, any method that yields the right answer in general will not be able to perform well for a problem with a large number of variables.

One way out of this situation is to utilize algebra to find *nonnegativity certificates*, i.e., conditions that are

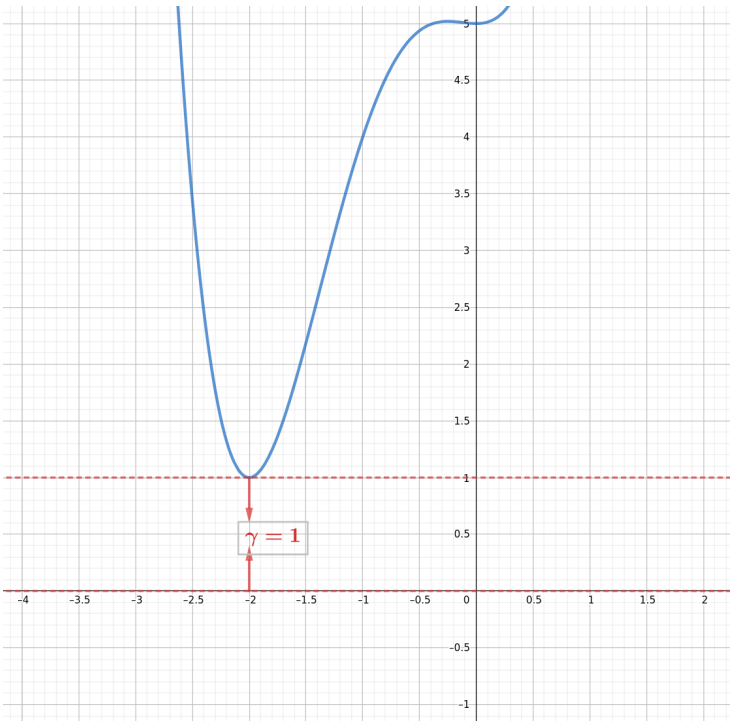

Figure 2.2: The red distance indicates the maximum amount of γ that one can subtract from $f(x) = x^4 + 3x^3 + x^2 + 5$ such that $f - \gamma$ is nonnegative.

1. easy to check,
2. imply nonnegativity,
3. are satisfied by a significantly large class of polynomials.

Besides its usefulness for polynomial optimization, the notion of nonnegativity on its own has been an attractive research area since late 19th century. A classical and well studied method to certify nonnegativity of a polynomial $f \in \mathbb{R}[\boldsymbol{x}]_{2d}$ is to write f as *sum of squares of polynomials*, i.e. find $s_1, \ldots, s_k \in\in \mathbb{R}[\boldsymbol{x}]_d$ such that $f = \sum_{i=1}^{k} s_i^2$. By the end of the 19th century, it was known that the a univariate polynomial is nonnegative if and only if it is a sum squares of polynomials. However, this statement does not hold for general multivariate polynomials of degree at least 4 as shown by Hilbert in 1888 [Hil88].

Theorem 2.2.7 ([Hil88])**.**

$$P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right) = \left\{ f \in \mathbb{R}[x_1, \ldots, x_n]_{2d} \;\middle|\; f = \sum_k s_k^2 \text{ for some } s_k \in \mathbb{R}[x_1, \ldots, x_n]_d \right\}$$

if and only if $(n, 2d) \in \{(k, 2), (2, 4), (1, k) \;:\; k \in \mathbb{N}_{>0}\}$.

Hilbert later posed a generalization of his earlier results as his 17th problem in his famous 1900 address to International Congress of Mathematicians in Paris [Hil00] [1], which was answered affirmatively by Emil Artin [Art27]. In Section 2.3, we provide a summary of the sums of squares techniques for global nonnegativity certification.

Another classical way to certify the nonnegativity of a polynomial is to use the *inequality of arithmetic and geometric means*, short *AM-GM inequality*, which states that:

$$\sum_{j=0}^{d} \lambda_j t_j - \prod_{j=0}^{d} t_j^{\lambda_j} \geq 0, \tag{2.2.4}$$

for $t_j \geq 0$, $\lambda_j \geq 0$ and $\sum_{j=1}^{d} \lambda_j = 1$, e.g., see [Ste10, Chapter 2]. In Section 2.4, we discuss how one can use the AM-GM inequality systematically to certify nonnegativity of polynomials. For now, we see how to use the AM-GM inequality for nonnegativity certification in a basic example.

Example 2.2.8. Let $t_1 = x^4, t_2 = y^4, t_3 = 1$ and $\lambda_1 = \lambda_2 = \frac{1}{4}$ and $\lambda_3 = \frac{1}{2}$, then the AM-GM inequality states that

$$x^4 + y^4 + 2 - 4xy \geq 0.$$

Alternatively, we can see that this expression is nonnegative since it is a sum of squares:

$$x^4 + y^4 + 2 - 4xy = (x^2 - 1)^2 + (y^2 - 1)^2 + 2(x - y)^2.$$

⬡

In Section 2.4 we cover the theory and history of this nonnegativity certificate, and provide some more interesting examples.

2.3 Cone of Sums of Squares of Polynomials

As we briefly discussed at the end of Section 2.2, the classical and most common certificate of nonnegativity are sums of squares (SOS). In this subsection we discuss sums of squares as certificates of nonnegativity, and introduce the SOS cone that will later be used in the thesis. A detailed theoretical overview of the subject can be found in e.g., [BPT12, Mar08, Las10].

[1] We note that this citation is a translation of Hilbert's original text. The original text in German can be found in `http://www.deutschestextarchiv.de/book/show/hilbert_mathematische_1900`

A polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]_{2d}$ is called a *sum of squares (SOS)* if

$$f(\boldsymbol{x}) = \sum_k s_k^2$$

for some $s_k \in \mathbb{R}[x_1, \ldots, x_n]_d$. For arbitrary fixed n and $2d$, we define the following subset of the full-dimensional convex cone $P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right)$:

$$\Sigma_{n,2d} \;:=\; \left\{f \in P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right) \;:\; f \text{ is SOS}\right\}.$$

We note here that dehomogenization and homogenization to an even degree preserves the property of being SOS. $\Sigma_{n,2d}$ is a cone, since adding two SOS polynomials and multiplying an SOS polynomial with a nonnegative scalar results in another SOS polynomial. The fact that $\Sigma_{n,2d}$ is convex also follows easily, since $tf + (1-t)g$ is an SOS for any $f, g \in \Sigma_{n,2d}$ and $t \in [0,1]$. Therefore, $\Sigma_{n,2d}$ is a cone inside the vector space $\mathbb{R}[x_1, \ldots, x_n]_{2d}$, and $\Sigma_{n,2d} \subseteq P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right)$ since each SOS is a priori nonnegative. We call the convex cone $\Sigma_{n,2d}$ the *SOS Cone* of n-variate polynomials with maximum degree $2d$. Furthermore, $\Sigma_{n,2d}$ is a closed cone as proven by Robinson [Rob69], and it is full dimensional for all n and $2d$.

However, as we mentioned before, SOS cone is strictly contained in cone of positive polynomials due to Theorem 2.2.7. In fact, the cone $P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right)$ is far larger than the cone $\Sigma_{n,2d}$ for large values of n and $2d$, as pointed out in [Ble06]. Though, it took a while for mathematicians to find a concrete example of a nonnegative polynomial that is not SOS. Hilbert pointed out a polynomial in $P\left(\mathbb{R}[x_1, x_2]_6\right)$ that is not SOS, as he proved Theorem 2.2.7. However, his proof in [Hil88] is considered to be nonconstructive, since he uses certain polynomials that are known to exists only theoretically. The first concrete example appeared almost 80 years later: In 1967 [Mot67], Motzkin pointed out, using the AM-GM inequality, that all polynomials of the form

$$\left(x_1^2 + \cdots + x_n^2 - n - 1\right) x_1^2 \cdots x_n^2 + 1 \tag{2.3.1}$$

are nonnegative for $n \geq 2$.

Example 2.3.1. We note that for $n = 2$, (2.3.1) corresponds to the famous *Motzkin polynomial*. If we let $t_1 = x^4y^2, t_2 = x^2y^4, t_3 = 1$ and $\lambda_1 = \lambda_2 = \lambda_3 \frac{1}{3}$, then the AM-GM inequality 2.2.4 implies that

$$M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2 \geq 0.$$

Unlike Example 2.2.8 and (2.4.1), the Motzkin polynomial cannot be represented as a sum

of squares of polynomials. The Motzkin polynomial is an important historical example, because it was the first concrete example of a nonnegative polynomial that is not contained in $\Sigma_{n,2d}$ for any n and d. ⬡

Showing that the Motzkin polynomial is not in $\Sigma_{2,6}$ requires a careful inspection of the coefficients. We describe a method to see why the Motzkin polynomial cannot be written as a sum of squares.

Proposition 2.3.2. The polynomial $M(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2 \in P(\mathbb{R}[x,y])$ is not in $\Sigma_{2,2d}$ for any d.

Proof. (Term Inspection Method) Let us name the coefficients of $M(x,y)$ as follows:

$$\sum_{\substack{i,j\in\mathbb{N},\\ i+j\leq 6}} \mathscr{c}_M(i,j)\cdot x^i y^j,$$

and assume for the sake of contradiction that $M(x,y) = \sum_k s_k(x,y)^2$ for some $s_k(x,y) \in \mathbb{R}[x,y]$. Since $M(x,y)$ is of degree 6, we can consider each $s_k(x,y)$ in the 10 dimensional vector space $\mathbb{R}[x,y]_3$, and denote each $s_k(x,y)$ as

$$\begin{aligned}\mathscr{c}_{s_k}(3,0)\cdot x^3 + \mathscr{c}_{s_k}(2,1)\cdot x^2y + \mathscr{c}_{s_k}(1,2)\cdot xy^2 + \mathscr{c}_{s_k}(0,3)\cdot y^3\\ +\mathscr{c}_{s_k}(2,0)\cdot x^2 + \mathscr{c}_{s_k}(1,1)\cdot xy + \mathscr{c}_{s_k}(0,2)\cdot y^2\\ +\mathscr{c}_{s_k}(1,0)\cdot x + \mathscr{c}_{s_k}(0,1)\cdot y\\ +\mathscr{c}_{s_k}(0,0).\end{aligned}$$

First, by comparing the value of the coefficient $\mathscr{c}_M(2,2)$ in the two representations, we have

$$\sum_k \left(\mathscr{c}_{s_k}(1,1)^2 + \mathscr{c}_{s_k}(2,1)\mathscr{c}_{s_k}(0,1) + \mathscr{c}_{s_k}(1,2)\mathscr{c}_{s_k}(1,0)\right) = -3. \tag{2.3.2}$$

Our aim is to find a contradiction by showing that $\mathscr{c}_{s_k}(0,1)$ and $\mathscr{c}_{s_k}(1,0)$ are zero.

Since the coefficient of x^6 is zero in $M(x,y)$, it easily follows that $\mathscr{c}_{s_k}(3,0) = 0$ for all k. Next, we consider the coefficient of x^4 in $M(x,y)$, which is on the one hand equal to zero, but on the other hand equal to $\sum_k (\mathscr{c}_{s_k}(2,0)^2 + 2\mathscr{c}_{s_k}(3,0)\mathscr{c}_{s_k}(1,0))$ in the SOS representation of $M(x,y)$. Since $\mathscr{c}_{s_k}(3,0) = 0$, we see that $\mathscr{c}_{s_k}(2,0) = 0$ for all k as well. Similarly, comparing the coefficient of x^2 yields the relation

$$0 = \sum_k \left(\mathscr{c}_{s_k}(1,0)^2 + 2\mathscr{c}_{s_k}(2,0)\mathscr{c}_{s_k}(0,0)\right).$$

Since $c_{s_k}(2,0) = 0$ for all k, the relation above implies that $c_{s_k}(1,0) = 0$ for all k. Note that $M(x,y)$ is symmetric in the variables, hence we can carry out the same argument by comparing the coefficients of y^6, y^4 and y^2, and show that $c_{s_k}(0,3) = c_{s_k}(0,2) = c_{s_k}(0,1) = 0$ for all k. Therefore, (2.3.2) reduces down to $\sum_k (c_{s_k}(1,1)^2) = -3$, which is a clear contradiction. So $M(x,y)$ cannot be written as a sum of squares of polynomials. □

The particular approach that we used in the proof of Proposition 2.3.2 was studied by Choi and Lam in [CL77b] and [CL77a], and named as *Term inspection method* by the authors. This approach was later generalized in [CLR95] by Choi, Lam and Reznick, and referred as the *Gram matrix method* in [Rez00].

If we want to use sum of squares as a nonnegativity certificate, then we need an efficient way to check if a polynomial f is in $\Sigma_{n,2d}$ or not. As it turns out, positive semi-definite matrices plays an important role in achieving this. This being the case, we now recall the definition of positive semi-definite matrices and some facts about them.

Definition 2.3.3. A matrix $A \in \mathbb{R}^{n\times n}$ is called *symmetric* if $A^T = A$, and A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is called *positive semi-definite (PSD)* if $\boldsymbol{x}^T A\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$. We denote the set of n by n symmetric matrices by $\mathcal{S}^n$, and positive semi-definite matrices by $\mathcal{S}^n_+$. ⬡

Note that $\mathcal{S}^n_+$ also forms a cone in its ambient vector space $\mathbb{R}^{n\times n}$. Indeed, if $A, B \in \mathcal{S}^n_+$ and $\alpha \in \mathbb{R}_{\geq 0}$, then $\alpha A + B$ is PSD since

$$\boldsymbol{x}^T (\alpha A + B)\, \boldsymbol{x} = \alpha \left(\boldsymbol{x}^T A\boldsymbol{x}\right) + \boldsymbol{x}^T B\boldsymbol{x} \geq 0$$

for any $\boldsymbol{x} \in \mathbb{R}^n$. In a similar manner, it is easy to see that $\mathcal{S}^n_+$ is a convex cone in $\mathbb{R}^{n\times n}$ which makes it quite convenient to study via convex optimization. There are several alternative characterizations of positive semi-definite matrices, and this means we have various methods to check if a given matrix is PSD. We mention some of these characterizations in Proposition 2.3.4, and we refer the reader to [BPT12, Appendix A] for a more inclusive list.

Proposition 2.3.4. Let $A \in \mathcal{S}^n$ be a symmetric matrix. Then the following statements are equivalent:

1. For all $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x}^T A\boldsymbol{x} \geq 0$.
2. A admits a *Cholesky decomposition*, i.e. there exists a factorization $A = BB^T$, where $B \in \mathbb{R}^{n\times r}$ and $\operatorname{rank}(A) = r$.
3. All eigenvalues of A are nonnegative.

4. All leading principal minors of A are nonnegative.

The Gram matrix method points out the connection between PSD matrices and SOS polynomials. Before we explain it, we have to set some more notation. Given $f \in \Sigma_{n,2d}$ with $\sum_{k=1}^{t} s_k^2$ for some $s_k \in \mathbb{R}[x_1, \ldots, x_n]_d$, we write $s_k = \sum_{\boldsymbol{\alpha} \in A_{s_k}} \mathscr{c}_{s_k}(\boldsymbol{\alpha}) \boldsymbol{x}^{\boldsymbol{\alpha}}$, where $\mathscr{c}_{s_k}(\boldsymbol{\alpha})$ denotes the coefficient of s_k at monomial $\boldsymbol{x}^{\boldsymbol{\alpha}}$. We define the following vector for each exponent $\boldsymbol{\alpha}$:

$$U_{\boldsymbol{\alpha}} := (\mathscr{c}_{s_1}(\boldsymbol{\alpha}), \ldots, \mathscr{c}_{s_t}(\boldsymbol{\alpha})) \in \mathbb{R}^t$$

for each exponent $\boldsymbol{\alpha} \in \mathbb{N}^n$ arising from the monomials in $\mathbb{R}[x_1, \ldots, x_n]_d$. For each $\boldsymbol{\alpha}, \boldsymbol{\alpha'} \in \mathbb{N}^n$ with $||\boldsymbol{\alpha}||_1 \leq d$ and $||\boldsymbol{\alpha'}||_1 \leq d$ we define

$$G(\boldsymbol{\alpha}, \boldsymbol{\alpha'}) \;:=\; B(\boldsymbol{\alpha}) \cdot B(\boldsymbol{\alpha'}) \;=\; \sum_{k=1}^{t} \mathscr{c}_{s_k}(\boldsymbol{\alpha}) \mathscr{c}_{s_k}(\boldsymbol{\alpha'}).$$

Definition 2.3.5. The $\binom{n+d}{n} \times \binom{n+d}{n}$ matrix

$$[G(\boldsymbol{\alpha},\ \boldsymbol{\alpha'})]_{\substack{\boldsymbol{\alpha}, \boldsymbol{\alpha'} \in \mathbb{N}^n \\ ||\boldsymbol{\alpha}||_1 \leq d, ||\boldsymbol{\alpha'}||_1 \leq d}}$$

is called *the Gram matrix* of f with respect to $s_1, \ldots, s_k \in \mathbb{R}[x_1, \ldots, x_n]_d$. ⬡

The Gram matrix method is based on the following observation given in [CLR95, Theorem 2.4], and in essence states that a polynomial is SOS if one can find a Gram matrix associated to it.

Theorem 2.3.6. [CLR95, Theorem 2.4] Let $f(\boldsymbol{x}) = \sum \mathscr{c}_f(\boldsymbol{\alpha}) \boldsymbol{x}^{\boldsymbol{\alpha}}$, and let

$$G = [G_{\boldsymbol{\alpha}, \boldsymbol{\alpha'}}]_{\substack{\boldsymbol{\alpha},\ \boldsymbol{\alpha'} \in \mathbb{N}^n \\ ||\boldsymbol{\alpha}||_1 \leq d, ||\boldsymbol{\alpha'}||_1 \leq d}}$$

be a $\binom{n+d}{n} \times \binom{n+d}{n}$ symmetric matrix. Then the following statements are equivalent:

- $f \in \Sigma_{n,2d}$ and G is the Gram matrix associated to f (with respect to some SOS representation $f = \sum_{k=1}^{t} s_k^2$ for some $s_k \in \mathbb{R}[x_1, \ldots, x_n]_d$);
- G is PSD and for all $\boldsymbol{\beta} \in \mathbb{N}^d$ with $||\boldsymbol{\beta}||_1 \leq 2d$

$$\mathscr{c}_f(\boldsymbol{\beta}) = \sum_{\substack{\boldsymbol{\alpha} + \boldsymbol{\alpha'} = \boldsymbol{\beta} \\ \boldsymbol{\alpha}, \boldsymbol{\alpha'}, \boldsymbol{\beta} \in \mathbb{N}^n}} G_{\boldsymbol{\alpha}, \boldsymbol{\alpha'}}.$$

Furthermore, the minimum number of squares required to represent f equals the minimal rank of all Gram matrices associated to f.

In [PW98], Powers and Wörmann implemented the Gram matrix method as an algorithm.

Let $f \in \mathbb{R}[x_1, \dots, x_n]_{2d}$ and $N = \binom{n+d}{n}$. Then according to Theorem 2.3.6 f admits a Gram matrix $G \in \mathcal{S}_+^N$ such that $f(\boldsymbol{x}) = \boldsymbol{v}^T G \boldsymbol{v}$ for $\boldsymbol{v} \in \mathbb{R}^N$ if and only if f is SOS. We find linear constraints on the entries of G by comparing the terms of the original f and the polynomial given by $\boldsymbol{v}^T G \boldsymbol{v}$. Let us now make an example on how to use Gram matrix method to find an SOS representation.

Example 2.3.7. Consider the polynomial $f(x_1, x_2) = 1 + 2x_1 + 4x_1x_2 + 5x_1^2 + x_2^2$, and let $\boldsymbol{v} := (1, x_1, x_2) \in \mathbb{R}^3$. If f admits a Gram matrix $G = (g_{ij})_{i,j\in[3]}$, then it must hold that

$$\begin{aligned} f(x_1, x_2) &= \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{12} & g_{22} & g_{23} \\ g_{13} & g_{23} & g_{33} \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \\ &= g_{11} + 2g_{12}x_1 + 2g_{13}x_2 + 2g_{22}x_1^2 + 2g_{33}x_2^2 + 2g_{23}x_1x_2. \end{aligned}$$

By comparing the coefficients we end up with the following conditions on the entries of G:

$$g_{11} = 1, \quad 2g_{12} = 2 \quad g_{13} = 0, \quad g_{22} = 5, \quad g_{33} = 1, \quad 2g_{23} = 4 \tag{2.3.3}$$

We see that the matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 5 & 2 \\ 0 & 2 & 1 \end{bmatrix}$ is a PSD matrix that satisfies all conditions in (2.3.3). Note that $A \in \mathcal{S}_+^3$ since A is symmetric with eigenvalues 0,1 and 6. If we let $\boldsymbol{v} = \begin{bmatrix} 1 & x_1 & x_2 \end{bmatrix}^T$, then

$$\begin{aligned} f(x_1, x_2) &= 1 + 2x_1 + 4x_1x_2 + 5x_1^2 + x_2^2 \\ &= \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 \\ 1 & 5 & 2 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \\ &= (1 + x_1)^2 + (2x_1 + x_2)^2 \end{aligned}$$

is in $\Sigma_{2,2}$. ⎔

Starting with an $G \in \mathcal{S}_+^n$, with the characterization given in Theorem 2.3.6 and using the Cholesky decomposition of G, one can write an SOS polynomial corresponding to any given choice of vector of monomials.

Example 2.3.8. Consider the Gram matrix G we calculated for the specific SOS representation of f given in Example 2.3.7. Similarly, if we decide to use $\boldsymbol{w} = \begin{bmatrix} x_1 & x_1x_2 & x_2 \end{bmatrix}^T$ as our vector of monomials, then we end up with a nonnegative polynomial $g(x_1, x_2) \in \Sigma_{2,4}$ as follows:

$$\begin{aligned} g(x_1, x_2) &= x_1^2 + x_1^2x_2 + 5x_1^2x_2^2 + 4x_1x_2^2 + x_2^2 \\ &= \begin{bmatrix} x_1 \\ x_1x_2 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 \\ 1 & 5 & 2 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_1x_2 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ x_1x_2 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_1x_2 \\ x_2 \end{bmatrix} \\ &= (x_1 + x_1x_2)^2 + (2x_1x_2 + x_2)^2. \end{aligned}$$

⎔

Theorem 2.3.6 gives a characterization of $\Sigma_{n,2d}$ in terms of positive semi-definite matrices. However, given an $f \in \mathbb{R}[x_1, \ldots, x_n]_{2d}$ how to find such a PSD matrix G, if there exists any, is not clear from statement of Theorem 2.3.6. In order to find such a PSD matrix, one can use semi-definite programming: as pointed out in [Par03, Theorem 3.3] by Parrilo, this is an SDP feasibility problem with matrix inequalities of size $\binom{n+2d}{n}$ by $\binom{n+2d}{n}$. A semi-definite program can be understood as a generalization of a linear program, where linear inequalities in the constraints are exchanged with linear matrix inequalities. A *semi-definite program*, or a *SDP*, is defined as the optimization problem:

$$\begin{aligned} &\text{minimize} && \operatorname{tr}(CX) \\ &\text{subject to} && \operatorname{tr}(A_iX) = b_i \\ & && X \text{ is PSD.} \end{aligned} \tag{2.3.4}$$

An important aspect of the problem defined in (2.3.4) is that the feasible set defined by the constraints is convex. We do not discuss SDPs in detail, however we point to the sources [VB96, WSV03, BPT12] for a comprehensive overview of the theory and the applications of SDPs. Yet, we note that the size of the SDP can be reduced especially if

the polynomial f is sparse. We also remark here that if $f \in H_{n,2d}$, then it is enough to consider only the monomial of degree d in $\boldsymbol{v}$, see Example 2.3.9.

Example 2.3.9. Let $f(x_1, x_2) = 2x_1^4 - x_1^2x_2^2 + 2x_1x_2^3 + 5x_2^4$, and define $\boldsymbol{v} = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix}$. If f admits a Gram matrix $G = (g_{ij})_{i,j\in[3]}$, then it must hold

$$\begin{aligned} f(x_1, x_2) &= \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix}^T \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{12} & g_{22} & g_{23} \\ g_{13} & g_{23} & g_{33} \end{bmatrix} \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix} \\ &= g_{11}x_1^4 + 2g_{13}x_1^3x_2 + (g_{33} + 2g_{12})x_1^2x_2^2 + 2g_{23}x_1x_2^3 + g_{22}x_2^4. \end{aligned}$$

Therefore, by comparing the coefficients we end up with the following conditions on the entries of G:

$$g_{11} = 2, \quad 2g_{13} = 0 \quad g_{33} + 2g_{12} = -1, \quad 2g_{23} = 2, \quad g_{22} = 5. \tag{2.3.5}$$

Then a positive semi definite G satisfying the linear equalities (2.3.5) can be found by solving the following semi-definite feasibility problem:

$$\begin{aligned} &\text{minimize} && 1 \\ &\text{subject to} && \operatorname{tr}(A_2X) = 0, \quad \operatorname{tr}(A_1X) = 2, \\ &&& \operatorname{tr}(A_3X) = -1, \quad \operatorname{tr}(A_4X) = 2, \\ &&& X \text{ is PSD}, \end{aligned}$$

with the matrices $A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$, $A_3 = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $A_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ and $A_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. A particular solution is given by:

$$G = \begin{bmatrix} 2 & -3 & 0 \\ -3 & 5 & 1 \\ 0 & 1 & 5 \end{bmatrix} = B^TB, \quad B = \begin{bmatrix} \frac{2}{\sqrt{2}} & 0 & 0 \\ \frac{-3}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{2}{\sqrt{2}} & \sqrt{3} \end{bmatrix},$$

and consequently we have the SOS representation:

$$f(x_1, x_2) = \left(\frac{2}{\sqrt{2}}x_1^2 - \frac{3}{\sqrt{2}}x_2^2\right)^2 + \left(\frac{1}{\sqrt{2}}x_2^2 + \sqrt{2}x_1x_2\right)^2 + \left(\sqrt{3}x_1x_2\right)^2.$$

⎔

We close this chapter by mentioning some significant directions, which are not part of this thesis, to tie up the loose ends of our discussion on the theory of SOS polynomials. There is a dual approach for polynomial optimization with SOS using the theory of moments, which was developed in parallel to Parrilo's approach independently by Laserre in [Las01]. For an extensive discussion of this dual approach, we refer the reader to [Las10]. As pointed out before, we are concerned with polynomial optimization in the unconstrained setting most of the time throughout this thesis. However, the theory we presented so far can be extended to the constrained case. This requires to work with the characterizations of nonnegativity over semi-algebraic subsets of $\mathbb{R}^n$, which are known as *Positivistellensätze* in the literature. The notion of a Positivstellensatz was first introduced by Krivine [Kri64] in 1964, and independently by Stengle [Ste74] in 1974. For compact semi-algebraic sets, the two most significant examples of Positivstellensätze are given by Schmüdgen [Sch91] and Putinar [Put93], see [BPT12, Section 3.4.3] and [Las10, Section 2.5].

2.4 Circuit Polynomials and SONC Cone

In this subsection we discuss an alternative nonnegativity certificate which is based upon another classical idea, namely the AM-GM inequality. We have already stated the AM-GM inequality in (2.2.4), and illustrated how to use it in a naive way as a nonnegativity certificate in Example 2.2.8 and Example 2.3.1. Now, we point out how to establish a general framework and make a more systematic approach using the AM-GM inequality.

The AM-GM inequality was a well understood fact by the end of 19th century, and throughout history many proofs of this fact have been written by mathematicians. In [HLP34, Page 17] authors even point out that the first nontrivial case of the inequality, i.e. when $d = 2$ and $\lambda_i = \frac{1}{2}$ for all i, can be proven using just two propositions from Euclid's elements. Among these many proofs, one of them was given by Adolf Hurwitz in 1891 [Hur91]. As it is alleged in [Rez00, Page 4], [Hur91] was the first published work that cited Hilbert's famous work [Hil88]. In this work, Hurwitz provides a new proof for

the AM-GM inequality, and shows that the polynomial

$$h(x_1, \dots, x_{2n}) = \sum_{j=1}^{2n} x_j^{2n} - 2n \prod_{j=1}^{2n} x_i \tag{2.4.1}$$

whose nonnegativity can be certified using the AM-GM inequality, is a sum of squares for all $n \in \mathbb{N}$. The forms given in (2.4.1) is an example that we revisit often, so we call them *Hurwitz forms* following Reznick's notation in [Rez89]. In [Hur91], Hurwitz mentions that (2.4.1) being SOS was not a trivial fact due to the 1888 result of Hilbert. As we have seen in the case of the Motzkin polynomial, not every nonnegative polynomial that arise from the AM-GM inequality is necessarily an SOS. We point out two more such examples that were studied by Choi and Lam in [CL77b] and [CL77a]:

$$Q(x, y, z, w) := x^2y^2 + x^2z^2 + y^2z^2 + w^4 - 4xyzw, \tag{2.4.2}$$

and

$$S(x, y, z) := x^4y^2 + y^4z^2 + z^4x^2 + -3x^2y^2z^2. \tag{2.4.3}$$

The nonnegativity of $Q(x, y, z, w)$ and $S(x, y, z)$ follows after applying the AM-GM inequality with a suitable choice of monomials, and it can be shown that they do not admit an SOS representation using a similar argument to the proof of Proposition 2.3.2.

In order to use the AM-GM inequality effectively as a nonnegativity certificate, we would like to work with a class of polynomials whose nonnegativity is implied by AM-GM, rather than individual examples like the Motzkin polynomial, Hurwitz forms, $Q(x, y, z, w)$ or $S(x, y, z)$. Before we introduce the main class of polynomials that we are interested in, we point out a historically important generalization of these examples.

In [Rez89], Reznick introduced a class of forms called *AGI-forms* by generalizing the examples of Hurwitz, Motzkin, Choi and Lam. We note that the Motzkin polynomial, Hurwitz forms, $Q(x, y, z, w)$ and $S(x, y, z)$ are all AGI-forms, and their supports share a property, which highlight the main aspect of an AGI-form: Their support contains only the vertices of the Newton polytope, plus one additional point in the interior. The name of the term is implied by AM-GM inequality, since the AGI-forms are constructed to be nonnegative using (2.2.4) with a suitable choice of monomial squares for t_j and $\lambda_j \in [0, 1]$ with $\sum \lambda_j = 1$. The first property means that the terms that correspond to the vertices in the Newton polytope cannot take negative sign, and this is a necessary condition for the nonnegativity of the polynomial, see e.g. [DIdW19, Proposition 2.1] or the first lemma in [Rez78, Page 365]. This fact will be useful especially in Chapter 4, where we will also prove a generalization of this fact, see Proposition 4.2.7. The second

property allows us to express the nonnegativity of the polynomial in a simple way with a single use of the AM-GM inequality. Reznick vastly studied the case of *simplicial AGI-forms*, i.e., the case where the Newton polytope of the AGI-form is a simplex, and gave a necessary and sufficient condition for a simplicial AGI-form to be SOS([Rez89, Corollary 4.9]). In Chapter 3, we further investigate how a polynomial f being SOS relates to the combinatorics of the support A_f, especially in the case of so called circuit polynomials which we introduce in Definition 2.4.1.

In recent years, Iliman and de Wolff [IdW16a] introduced circuit polynomials, which are generalizing the simplicial AGI-forms of Reznick. We note that every simplicial AGI-form is a nonnegative circuit polynomial, but in general AGI-forms are not circuit polynomials as they do not necessarily have simplex Newton polytope. Now, we proceed by giving the general definition of circuit polynomials, which is going to be the main object of study of this thesis.

Definition 2.4.1. A polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ is called a *circuit polynomial* if it is of the form

$$f(\boldsymbol{x}) = f_{\boldsymbol{\beta}}\boldsymbol{x}^{\boldsymbol{\beta}} + \sum_{j=0}^{r} f_{\boldsymbol{\alpha(j)}}\boldsymbol{x}^{\boldsymbol{\alpha(j)}} \tag{2.4.4}$$

for some $r \leq n$, exponents $\boldsymbol{\alpha(j)} \in 2\mathbb{N}^n, \boldsymbol{\beta} \in \mathbb{N}^n$, and the coefficients $f_{\boldsymbol{\alpha(j)}} \in \mathbb{R}_{>0}$, $f_{\boldsymbol{\beta}} \in \mathbb{R}$ such that $\mathrm{New}\,(f)$ is an r-dimensional simplex with vertices $\boldsymbol{\alpha(j)}$ and the exponent $\boldsymbol{\beta}$ is in the strict interior of $\mathrm{New}\,(f)$. We will sometimes refer to the exponent corresponding to $\boldsymbol{\beta}$ as the *inner term*. ⬡

The name "circuit" is inherited from the matroid theory, where a circuit means a minimal dependent set, see [Oxl11, Section 1.1]. The support of a given circuit polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ is a circuit, which is a minimal affine dependent set in $\mathbb{R}^n$ with $n+1$ vertices and one interior point. Each circuit polynomial comes with an associated circuit number, which is the key notion to study the nonnegativity of circuit polynomials.

Definition 2.4.2. Given a polynomial $f \in \mathbb{R}[x_1, \dots, x_n]$ as in Definition 2.4.1, let $\boldsymbol{\lambda} = (\lambda_0^{(\boldsymbol{\beta})}, \dots, \lambda_n^{(\boldsymbol{\beta})})$ denote the barycentric coordinates of $\boldsymbol{\beta}$ with respect to the $\boldsymbol{\alpha(j)}$s. Then, *the circuit number* Θ_f associated to the circuit polynomial $f(\boldsymbol{x})$ is

$$\Theta_p = \prod_{j=0}^{r} \left(\frac{f_{\boldsymbol{\alpha(j)}}}{\lambda_j^{(\beta)}} \right)^{\lambda_j^{(\boldsymbol{\beta})}}.$$

⬡

Given a circuit polynomial as in Definition 2.4.1, since $\boldsymbol{\alpha(0)}, \dots \boldsymbol{\alpha(k)}$ form the vertices of an r-dimensional simplex in $\mathbb{R}^n$, there exist unique barycentric coordinates of $\boldsymbol{\beta}$ with respect to the $\boldsymbol{\alpha(j)}$s. Therefore, the circuit number associated to a circuit polynomial f is indeed well-defined due to the unique barycentric representation of the inner term. Circuit polynomials are well suited to be used in nonnegativity certificates, because the nonnegativity of a circuit polynomial can be tested efficiently as proven in [IdW16a, Theorem 1.1].

Theorem 2.4.3. A circuit polynomial $f(\boldsymbol{x})$ given as in Definition 2.4.1 is nonnegative if and only if $|f_{\boldsymbol{\beta}}| \leq \Theta_f$ and $\boldsymbol{\beta} \notin (2\mathbb{N})^n$ or $f_{\boldsymbol{\beta}} \geq -\Theta_f$ and $\boldsymbol{\beta} \in (2\mathbb{N})^n$.

We note that this result has been proven by Fidalgo and Kovačec for circuit polynomials f with $\operatorname{New}(f) = \Delta_{2d}^n$ in 2010 in [FK10]. In order to prove Theorem 2.4.3 for a generic circuit polynomial, Iliman and de Wolff shows in [IdW16a, Proposition 3.1] it is enough to consider circuit polynomials whose Newton polytopes are scaled standard simplices.

All four of our previous examples, i.e., the Motzkin polynomial in Example 2.3.1, Hurwitz form in (2.4.1), $Q(x, y, z, w)$ in (2.4.2) and $S(x, y, z)$ in (2.4.3) are nonnegative circuit polynomials. Using Theorem 2.4.3 we can easily show this fact, see Example 2.4.4.

Example 2.4.4. As an example, consider $M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ from Example 2.3.1. If we let

$$\boldsymbol{\alpha(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \boldsymbol{\alpha(1)} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \ \boldsymbol{\alpha(2)} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \text{and } \boldsymbol{\beta} = \begin{bmatrix} 2 \\ 2 \end{bmatrix},$$

then we see that M is a circuit polynomial that is supported on the circuit given by the vertices $\boldsymbol{\alpha(i)}$ and the interior point $\boldsymbol{\beta}$. The barycentric coordinates of $\boldsymbol{\beta}$ with respect to the $\boldsymbol{\alpha(i)}$s are given as $\boldsymbol{\Lambda} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Thus, we calculate the circuit number:

$$\Theta_M = \left(\frac{1}{\frac{1}{3}}\right)^{\frac{1}{3}} \cdot \left(\frac{1}{\frac{1}{3}}\right)^{\frac{1}{3}} \cdot \left(\frac{1}{\frac{1}{3}}\right)^{\frac{1}{3}} = 3.$$

Since $\boldsymbol{\beta} \in (2\mathbb{Z})^n$ and $M_{\boldsymbol{\beta}} = 3 \geq 3 = -\Theta_M$, the polynomial $M(x, y)$ is a nonnegative circuit polynomial.

Similarly if we consider $Q(x, y, z, w) = x^2y^2 + x^2z^2 + y^2z^2 + w^4 - 4xyzw$ from (2.4.2), then by letting

$$\boldsymbol{\alpha(0)} = \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \ \boldsymbol{\alpha(1)} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 0 \end{bmatrix}, \ \boldsymbol{\alpha(2)} = \begin{bmatrix} 0 \\ 2 \\ 2 \\ 0 \end{bmatrix}, \ \boldsymbol{\alpha_{(3)}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 4 \end{bmatrix}, \text{ and } \boldsymbol{\beta} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

we see that $Q(x, y, z, w)$ is also a circuit polynomial that is supported on a 3-dimensional circuit. The barycentric coordinates of the inner term $\boldsymbol{\beta}$ is $\boldsymbol{\Lambda} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$, and the circuit number is

$$\Theta_Q = \left(\frac{1}{\frac{1}{4}}\right)^{\frac{1}{4}} \cdot \left(\frac{1}{\frac{1}{4}}\right)^{\frac{1}{4}} \cdot \left(\frac{1}{\frac{1}{4}}\right)^{\frac{1}{4}} \cdot \left(\frac{1}{\frac{1}{4}}\right)^{\frac{1}{4}} = 4.$$

Thus, again by Theorem 2.4.3 $Q(x, y, z, w)$ is nonnegative. ⬡

If a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]_{2d}$ is a *sum of nonnegative circuit polynomials(SONC)*, i.e., if $f = \sum_{k=1}^{t} s_k$ where each s_k is a nonnegative circuit polynomial, then f is clearly nonnegative.

Definition 2.4.5. Given a polynomial f and $A \subset \mathbb{N}^n$ such that $A_f \subset A$, we say that f admits a *SONC decomposition supported on A* if

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{\gamma} \in S} \lambda_{\boldsymbol{\gamma}} \boldsymbol{x}^{\boldsymbol{\gamma}} + \sum_{k=1}^{t} \lambda_k s_k(\boldsymbol{x}),$$

where $t \in \mathbb{N}$, each s_k is a nonnegative circuit polynomial with $A_{s_k} \subset A$, $S \subset 2\mathbb{N}^n \cap A_f$, $\lambda_{\boldsymbol{\gamma}}, \lambda_k \in \mathbb{R}_{>0}$ for all $k \in [t]$ and $\boldsymbol{\gamma} \in S$. In other words, a *SONC decomposition* of a polynomial f is a decomposition of f as a positive combination of monomial squares and nonnegative circuit polynomials. ⬡

In [IdW16a], *the cone of sums of nonnegative circuit polynomials* is initially introduced as the following subset of $\mathbb{R}[\boldsymbol{x}]_{2d}$:

$$\mathscr{C}_{n,2d} := \left\{ f = \sum_k \lambda_k g_k \in \mathbb{R}[x_1, \ldots, x_n]_{2d} \;\middle|\; \lambda_k \geq 0,\ g_k \text{ is a nonnegative circuit polynomial} \right\}. \tag{2.4.5}$$

Observe that $\mathscr{C}_{n,2d}$ is a convex cone, because for $a, b \in \mathbb{R}_{>0}$ and $f, g \in \mathscr{C}_{n,2d}$ it holds that $af + bg \in \mathscr{C}_{n,2d}$. Moreover, it was proven in [DIdW17, Theorem 4.3] $\mathscr{C}_{n,2d}$ is a full dimensional cone in $P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right)$. Recall that the dimension of $P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right)$ is $\binom{n+2d}{2d}$, i.e. it grows exponentially as n and d increase. This fact makes it a challenge even to represent the cone $\mathscr{C}_{n,2d}$ in digital environment, let alone to certify that a particular polynomial $f \in \mathscr{C}_{n,2d}$. To counteract this, we define an alternative SONC cone in Definition 2.4.8 which does not rely on considering all monomials in $\mathbb{R}[x_1, \ldots, x_n]_{2d}$. However, we first point out an important result that will motivate this next definition.

Theorem 2.4.6 ([W.20, Theorem 5.5],[MCW20a, Corollary 20])**.** Let f be a SONC polynomial with support A_f, then f admits a SONC decomposition where the support of each circuit polynomial in the sum is contained in A_f.

This key result means that a if f has a SONC decomposition, then it is always possible to find a SONC decomposition while preserving the sparsity of f. In contrast to the SONC case, an SOS representation of f usually contains more terms than there exist in the original support of f to make use of cancellation, see Example 2.4.7.

Example 2.4.7. Consider the polynomial $f(x,y) = 2x^4+2y^4+2-2xy^2-2x^2y$, which is an SOS polynomial. The initial support of f is $A_f = \{(0,0),(0,4),(4,0),(2,1),(1,2)\}$. In order to express f as an SOS, we require more monomials than the ones that arise from A_f. For example, one way to express f as a sum of squares of polynomials is as follows:

$$\left(x^2-1\right)^2+\left(y^2-1\right)^2+\left(x-y^2\right)^2+\left(x^2-y\right)^2$$

in which we had to introduce the new monomials x,y,x^2 and y^2 to the support. In fact, if we want to write f as an SOS, then one has to introduce new variables. For example, there is only one way that a term with the monomial x^4 can appear in such a sum: there has to be a term with the monomial x^2 in at least one of the summand. It is not clear which monomials should be added to the support of $f \in \Sigma_{n,2d}$ from the non-SOS representation of f, and typically one needs to consider all monomials in $\mathbb{R}[\boldsymbol{x}]_d$.

Alternatively, using circuit polynomials, we can verify the nonnegativity of f only working with A_f. We decompose f into circuit polynomials as

$$\begin{aligned} f(x,y) = 2x^4+2y^4+2-2xy^2-x^2y &= \left(x^4+y^4+1-2xy^2\right)+\left(x^4+y^4+1-2x^2y\right) \\ &= f_1(x,y)+f_2(x,y). \end{aligned}$$

Using Theorem 2.4.3, we see that f_1 and f_2 are nonnegative since $\Theta_{f_1} = 2\sqrt{2} \geq 2$ and $\Theta_{f_2} = 2\sqrt{2} \geq 2$. ⬡

As Theorem 2.4.6 suggest, it is enough to consider only those monomials in A_f in order to write down a SONC decomposition. This trait of SONC decompositions puts light to the fact that circuit polynomials are favorable for designing memory-friendly algorithms for nonnegativity certification. In view of this observation, we define a new cone for SONC polynomials whose support is contained in a given $A \subset \mathbb{N}^n$.

Definition 2.4.8. For a subset of lattice points $A \subset \mathbb{N}^n$, the *SONC cone over the support set A* is defined as

$$\mathscr{C}_A := \left\{ f \in \mathbb{R}^A \;\middle|\; f \text{ admits a SONC decomposition supported on } A \right\}.$$

⬡

We note that $\mathscr{C}_A$ is indeed a convex cone, since for any $a, b \in \mathbb{R}_{>0}$ and $f, g \in \mathscr{C}_A$ it holds that $af + bg \in \mathscr{C}_A$. It is clear that a polynomial f is nonnegative, if it is contained in $\mathscr{C}_A$. Moreover, as we will see in detail shortly, the containment of a polynomial $f \in \mathscr{C}_A$ can be formulated and efficiently solved as geometric programming problem.

Let us first formalize the notion of a geometric program. We call a function $\Psi : \mathbb{R}^n_{>0} \mapsto \mathbb{R}$ a *monomial function* if it is of the form

$$\Psi(x_1, \ldots, x_n) = c x_1^{\alpha_1} \cdots x_n^{\alpha_n},$$

where $c \in \mathbb{R}_{>0}$ and $\alpha_i \in \mathbb{R}$. A function $\Phi : \mathbb{R}^n_{>0} \mapsto \mathbb{R}$ is called a *posynomial function* if it is sum of monomial functions, i.e.,

$$\Phi(x_1, \ldots, x_n) = \sum_k c_k x_1^{\alpha_1^{(k)}} \cdots x_n^{\alpha_n^{(k)}},$$

where $c_k \in \mathbb{R}_{>0}$ and $\boldsymbol{\alpha^{(k)}} = \left(\alpha_1^{(k)}, \ldots \alpha_n^{(k)}\right) \in \mathbb{R}^n$.

Definition 2.4.9. A *geometric program*, or *GP*, is an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & \Phi_0(\boldsymbol{x}) \\ \text{subject to} & \Phi_1(\boldsymbol{x}), \ldots, \Phi_s(\boldsymbol{x}) \leq 1 \\ & \Psi_1(\boldsymbol{x}), \ldots, \Psi_t(\boldsymbol{x}) = 1 \end{array} \tag{2.4.6}$$

where $\Phi_1, \ldots, \Phi_s$ are posynomial functions, and $\Psi_1, \ldots, \Psi_t$ are monomial functions. ⬡

Geometric programs are convex, and can be solved with interior point methods. For a discussion on the computational complexity of this method, see [NN94, Section 6.3.1], and for further details about geometric programming we refer to [BV11, Chapter 4.5] and [BKVH07].

We note that the application of geometric programming to find global lower bounds for polynomials predates the definition of circuit polynomials. An important such application for our context is [GM12], where the authors explicitly employ geometric programming to find lower bounds for polynomials f with standard simplex Newton polytope,i.e., $\mathrm{New}(f) = \Delta^n_{2d}$. In particular, Ghasemi and Marshall point out an alternate sufficient condition for polynomials f with $\mathrm{New}(f) = \Delta^n_{2d}$ to be SOS in [GM12, Theorem 3.1]. Furthermore, the authors formulate a lower bound using their representation, and in [GM12, Corollary 3.6] write an explicit geometric program to compute this lower bound. In [IdW16b], the approach of Ghasemi and Marshall is generalized to use circuit polynomials effectively with geometric programming by Iliman and de Wolff. This approach

was initially studied through the notion of ST polynomials by Iliman and de Wolff in [IdW16b]. However, we cover a reformulation of Iliman an de Wolff's approach using Definition 2.4.8 for the sake of unifying the notation in this thesis.

Now, let us consider a polynomial $f \in \mathbb{R}[\boldsymbol{x}]_{2d}$ such that $\text{New}\,(f) = \{\mathbf{0}, \boldsymbol{\alpha(1)}, \ldots, \boldsymbol{\alpha(n)}\}$ is an n-dimensional simplex with $\text{Vert}\,(\text{New}\,(f)) \in 2\mathbb{N}^n$ and nonzero constant term. $\gamma \in \mathbb{R}$ is a lower bound for f, if it holds that $f - \gamma \in \mathscr{C}_{A_f}$. Let us denote the points in A_f which are not vertices of $\text{New}\,(f)$ with $\mathscr{D}(f)$ for now, i.e., $\mathscr{D}(f) := A_f \setminus \text{Vert}\,(\text{New}\,(f))$. Then we can express f as

$$f = f_0 + \sum_{k=1}^{n} f_{\boldsymbol{\alpha(k)}} \boldsymbol{x}^{\boldsymbol{\alpha(k)}} + \sum_{\beta \in \mathscr{D}(f)} f_\beta \boldsymbol{x}^{\boldsymbol{\beta}}. \tag{2.4.7}$$

Note that each $\beta \in \mathscr{D}(f)$ has a unique barycentric coordinate $\left(\lambda_0^{(\boldsymbol{\beta})}, \ldots, \lambda_n^{(\boldsymbol{\beta})}\right)$ with respect to $\boldsymbol{\alpha(k)}$s. We would like to write a nonnegative circuit polynomial for each $\boldsymbol{\beta} \in \mathscr{D}(f)$ with $f_{\boldsymbol{\beta}} < 0$, and the dimension of the Newton polytope of each circuit polynomial might vary with $\boldsymbol{\beta}$. This means, some $\lambda_k^{(\boldsymbol{\beta})}$ might be equal to zero, so we define $nz(\boldsymbol{\beta}) := \{k \mid \lambda_k^{(\boldsymbol{\beta})} \neq 0\}$. Now we set $A = \text{New}\,(f)$, and point out an alternative formulation from [IdW16b] for $f - \gamma$ to be in $\mathscr{C}_A$, which is more suitable for a geometric programming formulation.

Theorem 2.4.10. [IdW16b, Theorem 3.4] Let f be a polynomial given as in (2.4.7), and let $\gamma \in \mathbb{R}$. Assume that for every $(\boldsymbol{\beta}, k) \in \mathscr{D}(f) \times [n]$ there exists $a_{\boldsymbol{\beta},k}$ such that :

(1) If $\lambda_k^{(\boldsymbol{\beta})} > 0$, then $a_{\boldsymbol{\beta},k} > 0$,

(2) $|f_\beta| \leq \prod_{\substack{j \in nz(\boldsymbol{\beta}) \\ j \neq 0}} \left(\frac{a_{\boldsymbol{\beta},k}}{\lambda_k^{(\boldsymbol{\beta})}}\right)^{\lambda_k^{(\boldsymbol{\beta})}}$ for every $\boldsymbol{\beta} \in \mathscr{D}(f)$ with $\lambda_k^{(\boldsymbol{\beta})} = 0$,

(3) $f_{\boldsymbol{\alpha(k)}} \geq \sum_{\boldsymbol{\beta} \in \mathscr{D}(f)} a_{\boldsymbol{\beta},k}$ for all $k \in [n]$,

(4) $f_{\mathbf{0}} - \gamma \geq \sum_{\substack{\boldsymbol{\beta} \in \mathscr{D}(f) \\ \mathbf{0} \notin nz(\boldsymbol{\beta})}} \lambda_0^{(\boldsymbol{\beta})} |f_{\boldsymbol{\beta}}|^{1/\lambda_0^{(\boldsymbol{\beta})}} \prod_{\substack{j \in nz(\boldsymbol{\beta}) \\ j \neq 0}} \left(\frac{\lambda_k^{(\boldsymbol{\beta})}}{a_{\boldsymbol{\beta},k}}\right)^{\lambda_k^{(\boldsymbol{\beta})}/\lambda_0^{(\boldsymbol{\beta})}}$.

Then $f - \gamma$ is a sum of $|\mathscr{D}(f)|$ nonnegative circuit polynomials, whose Newton polytopes are faces of $\text{Vert}\,(\text{New}\,(f))$.

Given a polynomial f, if there is a γ such that there exist $a_{\boldsymbol{\beta},1}, \ldots, a_{\boldsymbol{\beta},n}$ which satisfy all the conditions given in Theorem 2.4.10, then γ is a lower bound for f over $\mathbb{R}^n$. We define the supremum of all $\gamma \in \mathbb{R}$ which satisfy the conditions of Theorem 2.4.10 as the

SONC lower bound of f, and we denote it by f^*_{SONC}. [IdW16b, Theorem 3.5] further points out that f^*_{SONC} coincides with the supremum of all $\gamma \in \mathbb{R}$ such that $f - \gamma = \sum_{j=0}^{t} s_j$ for some nonnegative circuit polynomials s_j where New (s_j) is a face of New (f). Now we are ready to state the problem of finding f^*_{SONC} as a geometric program.

Theorem 2.4.11. [IdW16b, Corollary 4.2] Let f be a polynomial given as in (2.4.7), and let R be the subset of an $n|\mathscr{D}(f)|$ given as

$$R := \{(a_{\boldsymbol{\beta},k}) \;\mid\; a_{\boldsymbol{\beta},k} \in \mathbb{R}_{>0} \text{ for all } (\boldsymbol{\beta}, k) \in \mathscr{D}(f) \times \{1, \dots, n\}\}.$$

Then $f^*_{SONC} = f_0 - \gamma^*$, where $\gamma^* \in \mathbb{R}$ is given as the output of the following geometric program:

$$\begin{aligned}
&\text{minimize} \quad \sum_{\substack{\boldsymbol{\beta} \in \mathscr{D}(f) \\ \mathbf{0} \notin nz(\boldsymbol{\beta})}} \lambda_0^{(\boldsymbol{\beta})} |f_{\boldsymbol{\beta}}|^{1/\lambda_0^{(\boldsymbol{\beta})}} \prod_{\substack{j \in nz(\boldsymbol{\beta}) \\ j \neq 0}} \left(\frac{\lambda_k^{(\boldsymbol{\beta})}}{a_{\boldsymbol{\beta},k}} \right)^{\lambda_k^{(\boldsymbol{\beta})}/\lambda_0^{(\boldsymbol{\beta})}} \quad \text{over the subset } R' \text{ of } R, \\
&\text{defined by} \quad (1) \sum_{\boldsymbol{\beta} \in \mathscr{D}(f)} \left(\frac{a_{\boldsymbol{\beta},,k}}{f_{\alpha_k}} \right) \leq 1 \text{ for every } 1 \leq k \leq n, \text{ and} \\
&\qquad\qquad |f_\beta| \prod_{k \in nz(\boldsymbol{\beta})} \left(\frac{\lambda_k^{(\boldsymbol{\beta})}}{a_{\boldsymbol{\beta},k}} \right) \leq 1 \text{ for all } \boldsymbol{\beta} \in \mathscr{D}(f) \text{ with } \lambda_0^{(\boldsymbol{\beta})} = 0.
\end{aligned}$$

It is possible to extend this SONC/GP approach to find lower bounds for polynomials with nonsimplex Newton polytope. Let $f \in \mathbb{R}[x_1, \dots, x_n]_{2d}$ be a polynomial supported on the set $A \subset \mathbb{N}^n$ such that Vert $(A) \in (2\mathbb{N})^n$ and the coefficient $f_\alpha > 0$ for all $\boldsymbol{\alpha} \in$ Vert (A), and note that A_f is not necessarily a simplex. Let

$$A_f^{\mathrm{MS}} := \{\boldsymbol{\alpha} \in A \;\mid\; \boldsymbol{\alpha} \in (2\mathbb{N})^n \text{ and } f_{\boldsymbol{\alpha}} > 0\}$$

denote the exponents that correspond to those terms of f which are monomial squares. Note that Vert $(A_f) \subset A_f^{\mathrm{MS}}$ always holds, but A_f^{MS} may contain some other exponents from $\mathscr{D}(f)$. In [DIdW19, Section 5], the authors took an approach by triangulating A_f^{MS}, and then invoking Theorem 2.4.11 on each cell of the triangulation. The authors point out in [DIdW19, Proposition 5.3] that we can find lower bounds for f by solving the GPs arising from the each cell of the triangulation. In Chapter 4, we employ a slightly different approach where we consider the covers of A rather than the triangulation.

Definition 2.4.12. We call a collection of simplicial basins $\Delta_1, \dots, \Delta_s$ *a cover of A* if

(1) $\mathrm{Vert}\,(\Delta_j) \subseteq A_f^{\mathrm{MS}}$ for all j such that $1 \leq j \leq s$,

(2) Each $\boldsymbol{\beta} \in A \setminus A_f^{\mathrm{MS}}$, $\boldsymbol{\beta} \in \Delta_j$ for some j such that $1 \leq j \leq s$.

⎔

Note that considering covers instead of triangulations gives us more flexibility, and essentially lets us consider more circuit polynomials to be used in a SONC certificate. This approach has been intensively studied in [SdW18], and has been implemented as a part of the polynomial software POEM [SdW19] which is available in the following link:

http://www.iaa.tu-bs.de/AppliedAlgebra/POEM/

We further note that a class of nonnegative functions that was introduced by Chandrasekaran and Shah [CS16], which is called sums of arithmetic geometric mean exponential(SAGE) functions and motivated by signomial programming, can be seen as another generalization of AGI-forms. A *signomial* is a sum of exponentials

$$f(\boldsymbol{x}) = \sum_{j=0}^{t} f_{\boldsymbol{\alpha(j)}} e^{\langle \boldsymbol{\alpha(j)}, \boldsymbol{x} \rangle},$$

where $f_{\boldsymbol{\alpha(1)}}, \ldots, f_{\boldsymbol{\alpha(t)}} \in \mathbb{R}$, and $\boldsymbol{x}, \boldsymbol{\alpha(1)}, \ldots, \boldsymbol{\alpha(t)} \in \mathbb{R}^n$. A signomial f is called an *arithmetic geometric mean exponential* if f has at most one negative coefficient, and we say that f is a *SAGE* function if it is sum of arithmetic geometric mean exponentials. The connection between the signomials and polynomials is given by the component wise exponentiation function

$$exp : \mathbb{R}^n \to \mathbb{R}^n_{>0}, \quad (x_1, \ldots, x_n) \mapsto (e^{x_1}, \ldots, e^{x_n}). \tag{2.4.8}$$

Hence each signomial $\sum_{j=0}^{t} f_{\boldsymbol{\alpha(j)}} e^{\langle \boldsymbol{\alpha(j)}, \boldsymbol{x} \rangle}$ is in a bijective correspondence with the polynomial $\sum_{j=0}^{t} f_{\boldsymbol{\alpha(j)}} \boldsymbol{x}^{\boldsymbol{\alpha(j)}}$ on $\mathbb{R}^n_{>0}$ via the map given in (2.4.8). Such polynomials that arise from SAGE functions are called *SAGE polynomials*, and a SAGE polynomial with at most one negative term is called an $AM - GM$-polynomial. Note that checking the nonnegativity a SAGE polynomial f corresponds to checking nonnegativity of f on the positive orthant, and it is sufficient to consider $\mathbb{R}^n_{>0}$ instead of $\mathbb{R}^n_{\geq 0}$ since $\mathbb{R}^n_{>0}$ is dense in $\mathbb{R}^n_{\geq 0}$. We note, without going into further detail, that SAGE polynomials also form a convex cone similarly to SONC polynomials, and the containment of a function in SAGE cone can be formulated as an relative entropy program. For more details on SAGE functions, see [CS16, MCW20a, MCW20b]. As pointed out in [W.20, Theorem 1.1] (see also [MCW20a, Theorem 4] and [FdW19, Theorem 4.4]) the cones described by circuit polynomials and arithmetic-geometric mean exponential functions are the same.

There are various aspects of SONC polynomials and SONC cones that we did not discuss in this section, mainly because most of these will not be relevant for the course of this thesis. Yet, we would like to close this section by mentioning some significant works, in order to give a more complete picture of the theory of circuit polynomials. One of the discussions we omitted was how to use SONC polynomials in the constrained optimization setting, but here we point out a series of works in this direction: see [DIdW19] for an initial discussion of the topic, [DIdW17] for a Schmüdgen type Positivstellensatz for SONC polynomials, [DKdW18] for SONC optimization over Boolean hypercube constraints. Furthermore, for a broad comparison of other nonnegativity certificates with SONC we refer reader to [KdW19]. Additionally, the dual of the SONC cone has been recently introduced in [DNT18]. The dual SONC cone is further studied in [KNT19], and has been applied to polynomial optimization in [DHNdW20].

Chapter 3

Classification of Maximal Mediated Sets

The aim of Chapter 3 is to systematically study the notion of maximal mediated sets, and to present the work that has been done in [HRdWY20]. In order to do so, we first introduce the basic definitions and facts on maximal mediated sets in Section 3.1. In Section 3.2, we define an equivalence relation on maximal mediated sets, and show that each class can be represented with a lattice. Next, in Section 3.3, we discuss the database which we constructed using the aforementioned equivalence relation.

3.1 Maximal Mediated Sets

Section 3.1 is dedicated to give a comprehensive introduction to the notions of mediated and maximal mediated sets, which are required to present the results of [HRdWY20]. This section is divided into two parts: Section 3.1.1 consists of general definitions and facts that were already stated in some work prior to [HRdWY20], or their reformulations according to our purposes. First, we introduce the notion of mediated and maximal mediated sets with a motivation to study the nonnegative circuit polynomials and SOS polynomials in a common framework. Furthermore, we present various examples of maximal mediated sets, some special classes of maximal mediated sets, and two algorithms to compute the maximal mediated sets. In Section 3.1.2, we discuss some basic facts about maximal mediated sets which are pointed out in [HRdWY20]. Although these facts are not required in the rest of the Chapter 3, we prefer mention them in this section. The main reason is that these facts not only relate the maximal mediated sets to a larger class of polynomials, but also further motivate our investigation of maximal mediated sets.

3.1.1 General Introduction to Maximal Mediated Sets

Maximal mediated sets arise naturally from the study of nonnegative polynomials supported on a circuit; see [IdW16a]. Given a circuit polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ as in Definition 2.4.1, $\operatorname{New}(f)$ is an r-dimensional simplex with $\operatorname{Vert}(\operatorname{New}(f)) \subset 2\mathbb{N}^n$. Accordingly, in a big portion of this chapter, we consider the maximal mediated sets of integral simplices with vertices in $(2\mathbb{Z})^n$. With this in mind, if $S \subset \mathbb{R}^n$ is a k-dimensional simplex with $\operatorname{Vert}(S) \subset 2\mathbb{N}^n$, we call $\Delta = \operatorname{Vert}(S)$ an *k-simplicial basin*. We define the *total degree of a given k-simplicial basin* Δ as the maximal one norm of vectors in $\operatorname{conv}(\Delta)$, and note that this corresponds to the maximal degree of a polynomial with Newton polytope Δ. In order to refer to some results of Reznick, following his notation we call a k-simplicial basin Δ a *trellis*, if all of its elements have the same 1-norm. Most of the time we consider full dimensional simplices, since we can always embed any k-dimensional Newton polytope in $\mathbb{R}^n$ into $\mathbb{R}^k$ by simple change of coordinates and projection. Therefore, unless it is stated otherwise, $\operatorname{conv}(\Delta) \subset \mathbb{R}^n$ will be an n-dimensional simplex with vertices in $2\mathbb{Z}^n$. The maximal mediated set associated to a simplicial basin Δ is a subset of lattice points in $\operatorname{conv}(\Delta)$. Our motivation to study the maximal mediated sets of this particular class of polytopes originates from a fact that was pointed out by Iliman and de Wolff in [IdW16a, Theorem 5.2], which states that a nonnegative circuit polynomial f is in $\Sigma_{n,2d}$ if and only if $\boldsymbol{\beta}$ is in the maximal mediated set of $\operatorname{New}(f)$.

Theorem 3.1.1 ([IdW16a], Theorem 5.2)**.** Let

$$f(\boldsymbol{x}) = f_{\boldsymbol{\beta}}\boldsymbol{x}^{\boldsymbol{\beta}} + \sum_{j=0}^{n} f_{\boldsymbol{\alpha(j)}}\boldsymbol{x}^{\boldsymbol{\alpha(j)}}$$

be a nonnegative circuit polynomial where $\Delta = \{\boldsymbol{\alpha_0}, \ldots, \boldsymbol{\alpha_n}\} \subset \mathbb{Z}^n$ is an n-simplicial basin, $\boldsymbol{\beta} \in \operatorname{conv}(\Delta) \cap 2\mathbb{Z}^n$, $f_{\boldsymbol{\alpha_k}} \in \mathbb{R}_{>0}$ and $f_{\boldsymbol{\beta}} \in \mathbb{R}$. Then, f is SOS if and only if $\boldsymbol{\beta} \in \Delta^*$ or $f_{\boldsymbol{\beta}} > 0$ and $\boldsymbol{\beta} \in 2\mathbb{N}^n$.

As a historical remark, we note that Theorem 3.1.1 was proven in [Rez89, Corollary 4.9] for the special case of simplicial AGI-forms. As nonnegative circuit polynomials are the building blocks of the SONC cone, by studying the maximal mediated sets of simplicial basins one can argue to what extent the SONC cone consists of SOS polynomials.

Definition 3.1.2. Let Δ be an n-simplicial basin in $\mathbb{R}^n$. Then a subset of lattice points $M \subset \mathbb{Z}^n$ is called *Δ-mediated* if,

(1) $\Delta \subset M$, and

(2) given $\boldsymbol{p} \in M \setminus \Delta$, then there exist $\boldsymbol{q_1}, \boldsymbol{q_2} \in (2\mathbb{Z})^n \cap M$ such that $\boldsymbol{q_1} \neq \boldsymbol{q_2}$ and $\boldsymbol{p} = \dfrac{1}{2}(\boldsymbol{q_1} + \boldsymbol{q_2})$.

⬡

For any n simplicial basin Δ in $\mathbb{R}^n$, there exists at least one $M \subset \mathbb{Z}^n$ that is Δ-mediated, e.g. Δ itself is Δ-mediated. In Proposition 3.1.3, we show that Δ is the smallest Δ-mediated set which is contained in every Δ-mediated set. This observation was pointed out in [HRdWY20], but we provide a a full proof here.

Proposition 3.1.3. Let Δ be an integral simplex with vertices Δ in $(2\mathbb{Z})^n$, and let $M \subset \mathbb{Z}^n$ be an Δ-mediated subset. Then, $\operatorname{conv}(M) \subset \operatorname{conv}(\Delta)$.

Proof. Let Δ and M be given as in the statement. If $\operatorname{conv}(M) \not\subset \operatorname{conv}(\Delta)$, then $\operatorname{conv}(M)$ has a vertex $\boldsymbol{m} \in \mathbb{Z}^n$ that is not contained in $\operatorname{conv}(\Delta)$. Since M is Δ-mediated, it has to satisfy the property (2) of Definition 3.1.2. However, $\boldsymbol{m}$ cannot be written the as midpoint of two distinct even points in M since it is a vertex of $\operatorname{conv}(M)$. □

Proposition 3.1.3 implies that any Δ-mediated set is a finite subset of lattice points in $\operatorname{conv}\Delta$. We give two examples and one non-example in Example 3.1.4.

Example 3.1.4. Consider the simplex $\Delta \subset \mathbb{R}^n$, and the sets of lattice points $M_1, M_2, M_3 \in \mathbb{Z}^n$ given as follows:

$$\begin{aligned}
\Delta &= \{(0,0),(4,0),(0,4)\},\\
M_1 &= \{(0,0),(4,0),(0,4),(0,2),(2,2),(1,2)\},\\
M_2 &= \{(0,0),(4,0),(0,4),(2,0),(1,2)\},\\
M_3 &= \{(0,0),(4,0),(0,4),(2,0),(0,2),(2,2),(2,3)\}.
\end{aligned}$$

Δ is clearly contained in M_1, M_2, and M_3 so the condition (1) on Definition 3.1.2 is satisfied for M_1, M_2, and M_3. Let us first see that each element in $M_1 \setminus \Delta$ satisfies the condition (2). The elements $(0,2)$ and $(2,2)$ satisfy the condition (2), since they are midpoints of Δ. Furthermore, $(1,2)$ is the midpoint of $(2,2),(0,2) \in (2\mathbb{Z})^n \cap M_1$. In order to see that M_2 Δ-mediated, all we have to check is that the point $(1,2)$ satisfies the condition (2) of Definition 3.1.2. This is true since $(1,2)$ is the midpoint of $(2,0)$ and $(0,4)$ which are elements of $M_2 \cap (2\mathbb{Z})^2$.

However, M_3 is not Δ-mediated. Because, although $(2,3) \in M_3 \setminus \Delta$, $(2,3)$ cannot be written as a midpoint of two even integral points from $\operatorname{conv}(M_3)$ since it is a vertex of $\operatorname{conv}(M_3)$. See Figure 3.1 for a picture of M_1 and M_2, and see the left panel of Figure 3.2 for a picture of M_3.

⬡

As we can see from M_1 and M_2 in Example 3.1.4, two Δ-mediated sets are not necessarily comparable with respect to inclusion. Also, the intersection of two given Δ-mediated sets does not have to be Δ-mediated as well, see Example 3.1.5.

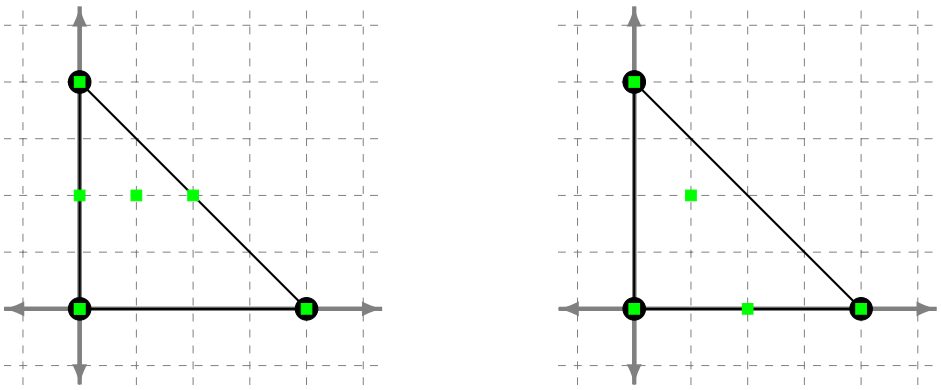

Figure 3.1: Two examples of Δ-mediated sets from Example 3.1.4: Black dots denote the simplicial basin Δ from Example 3.1.4, and green squares denote the sets M_1(left) and M_2(right), both of which are Δ-mediated.

Example 3.1.5. Let $\Delta \subset \mathbb{R}^2$ and $M_1, M_2 \subset 2\mathbb{N}^2$ be given as in Example 3.1.4, then

$$M_1 \cap M_2 = \{(0,0), (4,0), (0,4), (1,2)\}$$

We saw that M_1 and M_2 are Δ-mediated in Example 3.1.4. However, $M_1 \cap M_2$ is not Δ-mediated since $(1,2)$ is not a midpoint of two distinct points in $M_1 \cap M_2$. See the right panel in Figure 3.2 for an illustration. ⬡

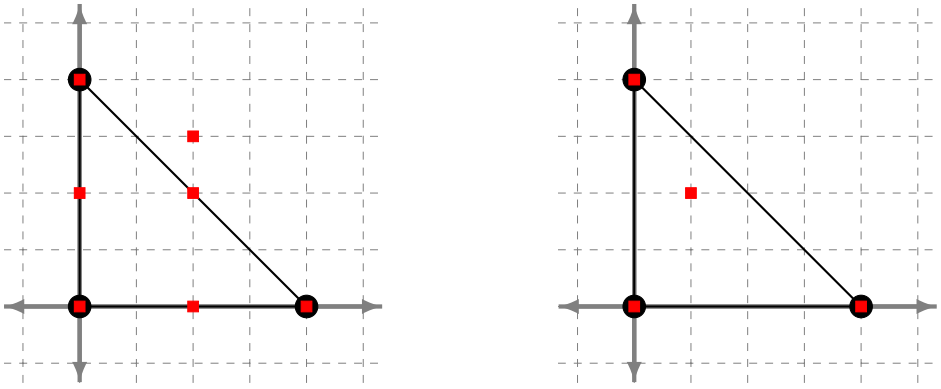

Figure 3.2: Black dots denote the simplicial basin Δ from Example 3.1.4, and red squares denote the sets M_3 from Example 3.1.4(left) and $M_1 \cap M_2$ from Example 3.1.5(right), both of which are not Δ-mediated.

Unlike intersection, the union of any two Δ-mediated sets for any n-simplicial basin $\Delta \subset \mathbb{R}^n$ is Δ-mediated.

Proposition 3.1.6. Let $\Delta \subset \mathbb{R}^n$ be a n-simplicial basin, and let M, N be two Δ-mediated sets. Then $M \cup N$ is also Δ-mediated.

We refer to condition (2) in Definition 3.1.2 repeatedly during our discussion on the maximal mediated sets. Therefore, before we provide an original proof of this basic observation, we introduce a notation to be able to refer condition (2) in Definition 3.1.2 easily.

Notation 3.1.7. Given $M \subseteq \mathbb{Z}^n$, by $\operatorname{Mid}(M)$ we denote the following set of *midpoints*:

$$\operatorname{Mid}(M) := \left\{ \frac{\boldsymbol{s}+\boldsymbol{t}}{2} : \boldsymbol{s}, \boldsymbol{t} \in M \cap (2\mathbb{Z})^n, \boldsymbol{s} \neq \boldsymbol{t} \right\}.$$

With this notation, M is Δ-mediated if and only if $\Delta \subset M$, and each $p \in M \setminus \Delta$ is in $\operatorname{Mid}(M)$.

Proof of Proposition 3.1.6. Let $\Delta \subset \mathbb{R}^n$ be an n-simplicial basin, and let M, N be two Δ-mediated sets. Since $\Delta \subset M$, condition (1) on Definition 3.1.2 is satisfied by $M \cup N$.

In order to show that condition (2) in Definition 3.1.2 holds for $M \cup N$, take an $\boldsymbol{p} \in (M \cup N) \setminus \Delta$, and without loss of generality, let $\boldsymbol{p} \in M \setminus \Delta$. Therefore, $\boldsymbol{p} \in \operatorname{Mid}(M)$ since M is Δ-mediated. Consequently, there exists $\boldsymbol{q_1}, \boldsymbol{q_2} \in M \cap 2\mathbb{N}^n$ such that $\boldsymbol{p} = \frac{1}{2}(\boldsymbol{q_1} + \boldsymbol{q_2})$. Clearly $\boldsymbol{q_1}, \boldsymbol{q_2} \in (M \cup N) \cap 2\mathbb{N}^n$, thus it follows that $\boldsymbol{p} \in \operatorname{Mid}(M \cup N)$. So, we conclude that $M \cup N$ is Δ-mediated. □

We give an immediate corollary of Proposition 3.1.6.

Corollary 3.1.8. Let $\Delta \subset \mathbb{R}^n$ be an n-simplicial basin. Then there exists a unique Δ-mediated set that contains all Δ-mediated sets.

Proof. Given an n-simplicial basin Δ, let $\mathscr{M}$ denote the set of all Δ-mediated sets in $\mathbb{R}^n$. $\mathscr{M}$ is not empty since $\Delta \in \mathscr{M}$. Consider the union

$$M^* := \bigcup_{M \in \mathscr{M}} M.$$

It follows from Proposition 3.1.6 that M^* is Δ-mediated. If there exists another such M_0 that contains all Δ-mediated sets, then it holds that $M_0 \subseteq M^*$ and $M^* \subseteq M_0$, i.e., $M_0 = M^*$. □

Based on Corollary 3.1.8, we give the following definition.

Definition 3.1.9. Given an n-simplicial basin Δ, the largest subset of $\mathbb{Z}^n$ that satisfies the two properties given in Definition 3.1.2 is called the *maximal mediated set* of Δ. We denote the maximal Δ-mediated set with Δ^*. ⬡

Remark 3.1.10. Given a nonnegative circuit polynomial as in Definition 2.4.1, *the MMS associated to f* is the maximal Δ-mediated set with $\Delta = \operatorname{Vert}(\operatorname{New}(f))$, denoted by $\Delta(f)^*$. ⬡

We note here that Reznick defined the maximal mediated sets in the context of trellises, see [Rez89]. The original statement of Reznick further points out a lower bound for the maximal mediated set of a given trellis. We reformulate Reznick's result in terms of our notation, and give an authentic proof using Proposition 3.1.6 and Corollary 3.1.8.

Theorem 3.1.11 ([Rez89], Theorem 2.2)**.** Given an n-simplicial basin $\Delta \subset \mathbb{R}^n$, there exists a unique maximal mediated set Δ^* satisfying

$$\Delta \cup \operatorname{Mid}(\Delta) \subseteq \Delta^* \subseteq \Delta \cap \mathbb{Z}^n.$$

Proof. Corollary 3.1.8 already shows that given an n-simplicial basin, Δ^* exists and it is unique. Note that we have already stated in Proposition 3.1.3 that any Δ-mediated set, in particular Δ^*, is a subset of $\operatorname{conv}(\Delta) \cap \mathbb{Z}^n$.

All that remains to show is that $\Delta \cup \operatorname{Mid}(\Delta) \subseteq \Delta^*$ for any n-simplicial basin Δ. It is clear that $\Delta \subset \Delta^*$ from property (1) of Definition 3.1.2. Let $\boldsymbol{p} \in \operatorname{Mid}(\Delta)$, and consider the set $M_{\boldsymbol{p}} := \{\boldsymbol{p}\} \cup \Delta$. For each $\boldsymbol{p} \in \operatorname{Mid}(\Delta)$, $M_{\boldsymbol{p}}$ is Δ-mediated as the only element $\boldsymbol{p}$ is in $M_{\boldsymbol{p}} \setminus \Delta$ is trivially in $\operatorname{Mid}(\Delta)$. Therefore, by definition Δ^* is a super set of all $M_{\boldsymbol{p}}$, and in particular it contains each $\boldsymbol{p} \in \operatorname{Mid}(\Delta)$. Thus, we conclude that $\Delta \cup \operatorname{Mid}(\Delta) \subseteq \Delta^*$. □

Note that in [Rez89], Reznick proved Theorem 3.1.11 in a more constructive manner using a different approach. Reznick does not only prove the existence of a unique maximal mediated set, but also provides an explicit algorithm to compute maximal mediated sets. We will omit the Reznick's full proof of Theorem 3.1.11, however we give his algorithm that constructs the Δ^* for a given Δ.

Algorithm 3.1.12 ([Rez89])**.** Given a finite $\Delta \subseteq (2\mathbb{Z})^n$, the following algorithm computes a non-increasing sequence of subsets that stabilizes at Δ^*.
Input: Δ: finite set of points in $(2\mathbb{Z})^n$
Output: Δ^*: the Δ-mediated subset of $\mathbb{Z}^n$ that contains every Δ-mediated set

1: $\Delta^0 \leftarrow \operatorname{conv}(\Delta) \cap \mathbb{Z}^n$
2: **repeat**
3: $\quad \Delta^i \leftarrow \operatorname{Mid}(\Delta^{i-1}) \cup \Delta$
4: **until** $\Delta^i = \Delta^{i-1}$
5: $\Delta^* \leftarrow \Delta^i$

Proof. The fact that the algorithm terminates, and the correctness of the algorithm follow from the proof of [Rez89, Theorem 2.2]. □

As pointed out in [IdW16b, Page 6], Reznick's construction in fact works for any set of even lattice points. However, for polynomials with non-simplicial Newton polytope, we are not aware of such an implication as in Theorem 3.1.1. In particular, Theorem 3.1.1 does not hold if New (f) is not a simplex, see Example 3.1.13 for a counterexample. As a consequence, we define the notion of being Δ-mediated only in the context of simplicial basins within this thesis.

Example 3.1.13. Consider the polynomial

$$p(x, y) = 2x^4y^2 + x^4 + 4x^2y^4 - 10x^2y^2 + 3,$$

and let $\Gamma = \text{Vert}\,(\text{New}\,(p)) = \{(0,0), (2,4), (4,2)(4,0)\}$. In fact, p is an AGI-form in Reznick's terms, and consequently it is nonnegative. Note that Γ is not a simplicial basin since $\text{conv}(\Gamma)$ is not a simplex in $\mathbb{R}^2$. If we run the Algorithm 3.1.12 with the input Γ, it returns that $\Gamma^* = \text{conv}(\Gamma) \cap \mathbb{Z}^2$. However, as it was pointed out in [IdW16a, Proposition 8.1], p is not SOS.

⬡

One can follow a slightly different approach than Algorithm 3.1.12 to compute the maximal mediated set of a simplicial basin Δ. Note that to compute Δ^*, it is enough to compute $\Delta^* \cap (2\mathbb{Z})^n$ since each $\boldsymbol{\alpha} \in \Delta^* \setminus (2\mathbb{Z})^n$ is midpoint of two distinct points in $\Delta^* \cap (2\mathbb{Z})^n$. With Algorithm 3.1.14, we compute $\Delta^* \cap (2\mathbb{Z})^n$ by starting with a lex-ordered list L of all points in $\text{conv}(\Delta) \cap (2\mathbb{Z})^n$ and iteratively removing all points that are not midpoints of two points in $\text{conv}(\Delta) \cap (2\mathbb{Z})^n$. Recall that given a finite lex-ordered a finite set $L \subseteq \mathbb{N}^n$, we denote the first and the last element of the lex ordered list L as $\text{head}(L)$ and $\text{tail}(L)$, respectively. We further note that, the following algorithm have been implemented as a class in `SAGE`[1] by Jacob Hartzer and Timo de Wolff, however it did not appear in the literature. Here we present our own proof that Algorithm 3.1.14 terminates and returns the correct result.

Algorithm 3.1.14. Given a finite $\Delta \subseteq (2\mathbb{Z})^n$, the following algorithm computes the maximal Δ-mediated set, Δ^*.
Input: Δ: finite set of affine independent points in $(2\mathbb{Z})^n$
Output: Δ^*: the Δ-mediated subset of $\mathbb{Z}^n$ that contains every Δ-mediated set

1: $L \leftarrow [(\text{conv}(\Delta) \cap (2\mathbb{Z})^n]$

[1]See http://www.iaa.tu-bs.de/timodewolff/MaximalMediatedSets.html for more information about this SAGE class

2: $i \leftarrow \text{tail}(L)$
3: **while** $i \neq \text{head}(L)$ **do**
4: **if** $i \notin \text{Mid}\,(L) \cup \Delta$ **then**
5: $L \leftarrow L \setminus \{i\}$
6: $i \leftarrow \text{tail}(L)$
7: **else**
8: decrement i
9: **end if**
10: **end while**
11: **return** $L \cup \text{Mid}\,(L)$

Proof. First we prove that the algorithm terminates, i.e., the while loop in the algorithm terminates. If L contains only one element, then the while loop immediately terminates since the condition in line 3 is satisfied. Assume L contains more than one element. If the condition in line 4 is not satisfied for any i as i runs through L, then the while loop terminates, because i is reduced in every run of line 8. If the condition in line 4 is satisfied for some i as i runs through L, then i is removed from L and while loop restarts. Since L has a finite cardinality k, the while loop terminates after at most $k-1$ restarts.

For the correctness of the algorithm, let Δ_0^* denote the output of Algorithm 3.1.12 with the input Δ. We show that

1. $\Delta^* \subseteq \Delta_0^*$, and
2. Δ_0^* is Δ-mediated

Thus, it follows that $\Delta^* = \Delta_0^*$ by Corollary 3.1.8 and the maximality of Δ^*.

As Δ^* is Δ-mediated, it satisfies the conditions given in Definition 3.1.2. If $\boldsymbol{p} \in \Delta^*$ is not an even point, then we can consider the points $\boldsymbol{q_1}, \boldsymbol{q_2} \in \Delta^* \cap (2\mathbb{Z})^n$ given as in the second property of Definition 3.1.2 instead of $\boldsymbol{p}$. Therefore, it is enough to show our first claim above only for even points, i.e. $\Delta^* \cap (2\mathbb{Z})^n \subset \Delta_0^* \cap (2\mathbb{Z})^n$. To argue by contradiction, we assume that $D = (\Delta^* \cap (2\mathbb{Z})^n) \setminus (\Delta_0^* \cap (2\mathbb{Z})^n)$ is not empty. When the algorithm is initialized, the list L is set to $\text{conv}(\Delta) \cap (2\mathbb{Z})^n$ which contains D. As the algorithm runs through, the elements of D are discarded one by one. Because otherwise, if some element of $\boldsymbol{d} \in D$ is not discarded when algorithm terminates, then it means that $\boldsymbol{d} \in \Delta_0^*$, which is a contradiction. Let $\boldsymbol{\alpha}$ denote the first element of D that is discarded from L. Note that $D \cap \Delta = \emptyset$ because Δ is a subset of both Δ^* and Δ_0^*. Let $L_{\boldsymbol{\alpha}}$ denote the elements that stay in the list L until $\boldsymbol{\alpha}$ is discarded. Since $\boldsymbol{\alpha} \in \Delta^* \setminus \Delta$, there exist distinct $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2} \in \Delta^*$ such that $\boldsymbol{\alpha} = \frac{\boldsymbol{\alpha_1}+\boldsymbol{\alpha_2}}{2}$. Since $\boldsymbol{\alpha}$ is assumed to be the first element discarded from L, both $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$ are in $L_{\boldsymbol{\alpha}}$. Because otherwise $\boldsymbol{\alpha_1}$ or $\boldsymbol{\alpha_2}$ would be the first element in D to be removed from L. Therefore, we have that $\boldsymbol{\alpha} \in \text{Mid}\,(L_\alpha)$ and the condition in step 4 fails

to hold. However, this implies that $\boldsymbol{\alpha}$ is not discarded which contradicts the fact that $\boldsymbol{\alpha} \in D$. Therefore, D is empty and $(\Delta^* \cap (2\mathbb{Z})^n) \subseteq (\Delta_0^* \cap (2\mathbb{Z})^n)$.

Last, in order to prove Δ_0^* is Δ-mediated, we let $\boldsymbol{\beta} \in \Delta_0^* \setminus \Delta$ and show that $\boldsymbol{\beta} \in \operatorname{Mid}(\Delta_0^*)$. Due to step 4, $\boldsymbol{\beta} \in \Delta_0^*$ only if $\boldsymbol{\beta} \in \operatorname{Mid}(\Delta^0 \cap (2\mathbb{Z})^n)$. The claim follows since $\operatorname{Mid}(\Delta_0^*) = \operatorname{Mid}(\Delta_0^* \cap (2\mathbb{Z})^n)$. □

Remark 3.1.15. The Algorithm 3.1.14 is implemented in POLYMAKE as an extension, see Remark 3.3.3. We see the further details of this implementation in Section 3.3. ⬡

We distinguish the simplicial basins that attain one of the two bounds given in Theorem 3.1.11. Following Reznick's notation, we call a simplicial basin Δ an *M-simplex* if $\Delta^* = \Delta \cup \operatorname{Mid}(\Delta)$, and an *H-simplex* if $\Delta^* = \operatorname{conv}(\Delta) \cap \mathbb{Z}^n$. We motivate this choice of the particular notation in Example 3.1.16.

Example 3.1.16. Consider the two simplicial basins $\Delta_1 = \{(0,0),(2,4),(4,2)\}$ and $\Delta_2 = \{(0,0),(4,0),(0,4)\}$ in $\mathbb{R}^2$. Following Algorithm 3.1.12, we compute that:

$$\Delta_1^* \;=\; \{(0,0),(1,2),(2,1),(2,4),(3,3),(4,2)\} = \Delta_1 \cup \operatorname{Mid}(\Delta_1)\,,$$

and

$$\Delta_2^* \;=\; \operatorname{conv}(\Delta_2) \cap \mathbb{Z}^2.$$

Δ_1 is an example of an M-simplex, and arises from the simplicial basin associated to the *Motzkin polynomial* which we defined in Example 2.3.1. In addition to its historic importance, it is the unique M-simplex among the 2-simplicial basin with maximal degree 6, see Remark 3.3.3. Δ_2 is an example of H-simplex. It arises from a factor-2 scaling of the simplicial basin associated to the *Hurwitz form* given in (2.4.1) where $2n = 2$. See Figure 3.3 for the visualization of Δ_1^* and Δ_2^*. ⬡

The case of 2-simplicial basins is less exciting than the higher dimensions, because most of the simplicial basins are actually H-simplices in this case.

Theorem 3.1.17. [IdW16a, Corollary 5.10] Let $\Delta \subset 2\mathbb{Z}^2$ be a 2-simplicial basin. If $\frac{1}{2}\operatorname{conv}\Delta$ contains at least four boundary lattice points then Δ is an H-simplex.

We note that the converse of Theorem 3.1.17 does not hold, see Example 3.1.18 for a counterexample.

Example 3.1.18. We consider the 2-simplicial basin $\Delta = \{(0,0),(2,6),(8,2)\}$, which is an H-simplex as can be verified with our `POLYMAKE` extension for MMS (see Remark 3.3.3). However, $\frac{1}{2}\operatorname{conv}(\Delta)$ contains only 3 integral points in its boundary. ⬡

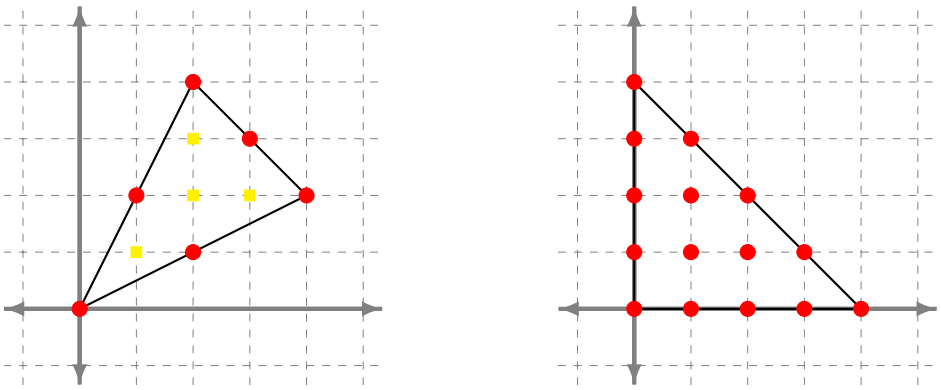

Figure 3.3: Δ_1(left) and Δ_2(right) in Example 3.1.16, red dots indicate the points that are in the MMS and yellow squares indicate the points that are not in the MMS.

For 2-simplicial basins, Reznick stated that they are always an M-simplex or an H-simplex; [Rez89, Page 9]. In this 1989 article, Reznick announced a proof for this claim, and another important result [Rez89, Proposition 2.7], but the particular article was not finished. For this reason, we state this claim as a conjecture in here.

Conjecture 3.1.19 (Page 9, [Rez89])**.** Let $\Delta \subset (2\mathbb{Z})^2$ be a simplicial basin, then Δ is either an M-simplex or an H-simplex.

We present an experimental result in Section 3.3 confirming this conjecture with simplicial basins of maximal 1-norm less than 150.

For a general n-simplicial basin Δ, the MMS does not necessarily attain one of the bounds given in Theorem 3.1.11, as we see in Example 3.1.20.

Example 3.1.20. Let $\Delta = \{(0,0,0,0),(0,0,0,4),(0,2,2,0),(2,0,2,0),(2,2,0,0)\}$. The convex hull of Δ contains 22 integral lattice points. Only two of these integral lattice points are not in Δ^*, namely $(1,1,1,0)$ and $(1,1,1,1)$. One can verify this via our software package discussed in Section 3.3. ⬡

We would like to point out some known necessary or sufficient conditions for a simplicial basin to be an M-simplex and H-simplex. Note that we already pointed out one such condition in Theorem 3.1.17. Another such condition was given in [Rez89, Proposition 2.7], referring to the same unfinished article along with Conjecture 3.1.19.

Theorem 3.1.21 ([Rez89, Proposition 2.7],[PW98, Theorem 3.1])**.** Given an n-simplicial basin Δ, k-factor scaling of Δ, i.e., $k\Delta$ is an H-simplex for every integer $k \geq \max 2, n-2$.

Very recently, Powers and Reznick proved [Rez89, Proposition 2.7] in [PR20]. However, after consulting with the authors we reached a consensus that their results in [PR20] do not solve Conjecture 3.1.19.

Example 3.1.22. Let Δ be as given in Example 3.1.20. Then,

$$2\Delta = \{(0,0,0,0),(0,0,0,8),(0,4,4,0),(4,0,4,0),(4,4,0,0)\}$$

is an H simplex due to Theorem 3.1.21. ⎔

The property of being an H-simplex is tied to some very well studied notions such as normality of a polytope, in our case the convex hulls of simplicial basins, see [IdW16a, Theorem 5.9].

Example 3.1.23. Let $\Delta = \{(0,0,0),(4,0,0),(0,6,0),(0,0,10)\} \subset (2\mathbb{Z})^3$. In [BG99, Example 2.2], Bruns and Gubeladze point out that the lattice polytope $\frac{1}{2}\operatorname{conv}(\Delta)$ is not normal. Indeed, $\frac{1}{2}\operatorname{conv}(\Delta)$ is not 2-normal due to the point $\boldsymbol{p} = (1,2,4)$, i.e., there exists no $\boldsymbol{p_1}, \boldsymbol{p_2} \in \mathbb{N}^3 \cap \operatorname{conv}(\Delta)$ such that $p_1 + p_2 = p$. Consequently, Theorem 5.9 of [IdW16a] implies that Δ cannot be an H-simplex. Indeed, $\boldsymbol{p} = (1,2,4)$ is the only point that is not in the MMS:

$$\Delta^* \;=\; \operatorname{conv}(\Delta) - \{(1,2,4)\}\,.$$

We visualize Δ^* in Figure 3.4. ⎔

In a like manner, there is a connection between M-simplices and *distict pair-sum (dps) polytopes* which are defined and studied in [CLR02].

Example 3.1.24. Let $\Delta = \{(0,0,0),(0,2,2),(2,0,2),(2,2,0)\}$, then Δ^* attains the lower bound in Theorem 3.1.11

$$\Delta^* \;=\; \operatorname{conv}(\Delta) \cap (\mathbb{Z}^n) - \{(1,1,1)\} \;=\; \Delta \cup \operatorname{Mid}(\Delta)\,.$$

One can verify that Δ is an M-simplex also using a result by Bommel [Bom14, Theorem 3.6] and the fact that $\frac{1}{2}\Delta$ is a distinct pair-sum (dps) polytope. Δ is the Newton polytope of

$$1 + x_1^2x_2^2 + x_1^2x_3^2 + x_2^2x_3^2 - 4x_1x_2x_3,$$

another well-known small example of a nonnegative polynomial that is cannot be written as sum of squares. In fact, this polynomial is the dehomogenized version of (2.4.2) ⎔

In Section 3.2, we introduce the necessary tools to measure how close a simplicial basin is to being an M-simplex or an H-simplex. The contents of Section 3.2 are required to understand the construction and the analysis of the MMS database given in [HRdWY20], see also Remark 3.3.4. In Section 3.3, we discuss the construction and some statistics

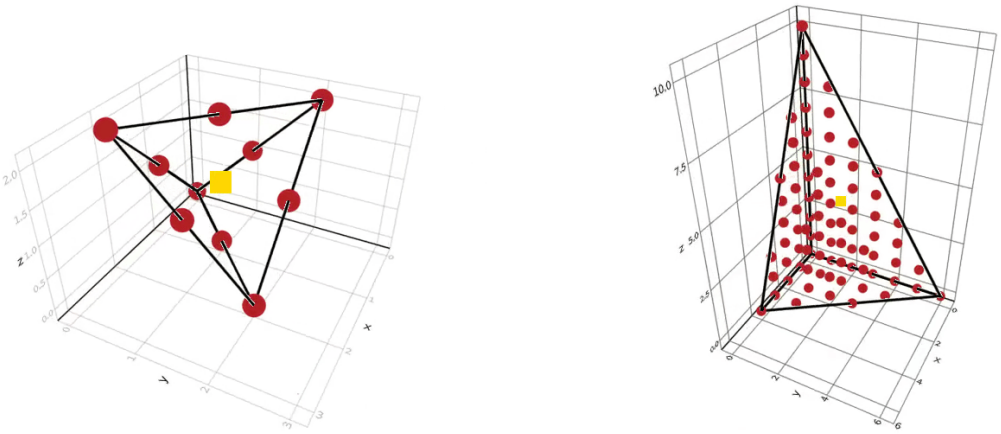

Figure 3.4: MMS of the simplicial sets given in Example 3.1.24(left) and Example 3.1.23(right), red dots indicate the points that are in the MMS and yellow squares indicate the points that are not in the MMS.

of our database. This discussion exposes the true variety of MMS in existence, and the fact that there are many simplicial basins that are not an M-simplex or an H-simplex. We close this section by pointing out some significant properties of the examples that we covered in this section which we verified using the database.

Remark 3.1.25. Using our implementation that was pointed out in Remark 3.1.15, we verified that Example 3.1.24 is the unique 3-simplicial basin with maximal degree 4 that attains the lower bound in Theorem 3.1.11. Furthermore, Example 3.1.20 is the only 4-simplicial basin with maximal degree at most 4 such that Δ^* is strictly between the bounds up to coordinate permutations. ⎔

3.1.2 A Generalization for SONC with Simplex Newton Polytope

In this section, we discuss a generalization of Theorem 3.1.1 to sums of nonnegative circuit polynomials with simplex Newton polytope. Recall that, our motivation to study the MMS was the characterization of circuit polynomials given by Theorem 3.1.1 according to their inner terms. In Theorem 3.1.26, we reinforce our motivation by showing that a characterization similar to Theorem 3.1.1 holds for SONC polynomials with simplex Newton polytope. As a result of this theorem, we point out Corollary 3.1.27, which we shows that the converse implication also holds in the part (2) of [IdW16b, Corollary 3.6]. The next proof heavily relies on the Gram matrix method; see e.g., Section 2.3 or [Rez00, Section 5.b] for an overview of the method. For $f \in \mathbb{R}[\boldsymbol{x}]_d$ and $\boldsymbol{\alpha} \in \mathbb{N}^n$, recall that $f_{\boldsymbol{\alpha}}$

denotes the coefficient of the term $\boldsymbol{x}^{\boldsymbol{\alpha}}$ in f.

Theorem 3.1.26. Let $\Delta = \{\mathbf{0}, \boldsymbol{\alpha}(\mathbf{1}), \ldots, \boldsymbol{\alpha}(\boldsymbol{n})\} \subset (2\mathbb{Z})^n$ be an n-simplicial basin, and let $Y = \{\boldsymbol{\beta}_\mathbf{1}, \ldots, \boldsymbol{\beta}_{\boldsymbol{m}}\} \subseteq \operatorname{int}(\operatorname{conv}(\Delta) \cap \mathbb{Z}^n)$ be a set of points. Let $f = \lambda_0 + \sum_{i=0}^{n} a_i \boldsymbol{x}^{\boldsymbol{\alpha}(\boldsymbol{i})} + \sum_{\boldsymbol{\beta} \in Y} b_{\boldsymbol{\beta}} \boldsymbol{x}^{\boldsymbol{\beta}}$ be a SONC with support $\Delta \cup Y$, $a_0, \ldots, a_n > 0$, such that for all $\boldsymbol{\beta} \in Y$, $b_{\boldsymbol{\beta}} < 0$ or $\boldsymbol{\beta} \notin (2\mathbb{Z})^n$. Then f is a sum of squares if and only if every $\boldsymbol{\beta} \in Y$ satisfies $\boldsymbol{\beta} \in \Delta^*$.

This theorem generalizes Theorem 3.1.1, which states the same result for the case $\#Y = 1$. Furthermore, this theorem links up the maximal mediated sets with a larger set of polynomials.

Proof. First, assume that f admits a SONC decomposition $f = \sum_{\boldsymbol{\beta} \in Y} s_{\boldsymbol{\beta}}$ where $s_{\boldsymbol{\beta}}$ is a nonnegative circuit polynomial with the support $\Delta \cup \{\boldsymbol{\beta}\}$. Since $s_{\boldsymbol{\beta}}$ is a nonnegative circuit polynomial satisfying $\Delta(s_{\boldsymbol{\beta}})^* = \Delta(f)^* = \Delta^*$, and since we have $\boldsymbol{\beta} \in \Delta^*$ by assumption, [IdW16a, Theorem 5.2] implies that $s_{\boldsymbol{\beta}} \in \Sigma_{n,2d}$. Thus, it follows that $f \in \Sigma_{n,2d}$.

For the converse, assume that $f \in \Sigma_{n,2d}$. We claim that given a $\boldsymbol{\beta} \in Y$, if

$$\boldsymbol{\beta} \notin (2\mathbb{Z})^n \text{ or } b_{\boldsymbol{\beta}} \leq 0 \tag{3.1.1}$$

then $\boldsymbol{\beta} \in \Delta^*$. Since $f \in \Sigma_{n,2d}$, it has a SOS decomposition $f = \sum_{i=i}^{k} h_i^2$. Define the set

$$M \;=\; \left\{\boldsymbol{\gamma} \in \mathbb{N}_d^n \;:\; \text{there exists an } i \in [k] \text{ with } h_{i\boldsymbol{\gamma}} \neq 0\right\},$$

where $h_{i\boldsymbol{\gamma}}$ denotes the coefficient of h_i at exponent $\boldsymbol{\gamma}$. For every $\boldsymbol{\beta} \in Y$ we define the set

$$L_{\boldsymbol{\beta}} \;=\; 2M \cup \Delta \cup \boldsymbol{\beta}.$$

We can assume $b_{\boldsymbol{\beta}} < 0$: if $\boldsymbol{\beta} \in (2\mathbb{Z})^n$, then $b_{\boldsymbol{\beta}} < 0$ by (3.1.1). Assume that $b_{\boldsymbol{\beta}} > 0$ and that there exists an odd entry β_j of $\boldsymbol{\beta}$ for some $j \in [n]$. After a transformation $\tau_j : x_j \mapsto -x_j$ we can consider $b_{\boldsymbol{\beta}} < 0$, while τ_j leaves $\sum_{i=1}^{k} h_i^2$ invariant; see also [IdW16a, proof of Theorem 5.2] and e.g., [BPT12]. With $b_{\boldsymbol{\beta}} < 0$ we obtain that $L_{\boldsymbol{\beta}}$ is Δ-mediated following verbatim the first part of the proof of [IdW16a, Theorem 5.2]. This completes the proof since every Δ-mediated set is contained in Δ^*.

□

Given $f = \lambda_0 + \sum_{i=0}^{n} a_i \boldsymbol{x}^{\boldsymbol{\alpha}(\boldsymbol{i})} + \sum_{\boldsymbol{\beta} \in Y} b_\beta \boldsymbol{x}^{\boldsymbol{\beta}} \in \mathbb{R}[x_1, \ldots, x_n]_{2d}$ such that $a_i > 0$, $b_{\boldsymbol{\beta}} < 0$ and $\operatorname{New}(f)$ is a simplex, due to [IdW16a, Theorem 5.5] we know that

$$f \in P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right) \iff f \text{ is SONC.}$$

The following corollary concerns two relaxations in polynomial optimization. First, we recall the following two quantities:

$$\begin{aligned} f_{\text{SOS}} &:= \max\{\lambda \in \mathbb{R} \mid f - \lambda \text{ is SOS}\} \\ f_{\text{SONC}} &:= \max\{\lambda \in \mathbb{R} \mid f - \lambda \text{ is SONC}\}. \end{aligned}$$

Both f_{SOS} and f_{SONC} are lower bounds for $f^* := \min\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^n\}$. Next, we present a corollary of Theorem 3.1.26 which generalizes the part (2) of [IdW16b, Corollary 3.6].

Corollary 3.1.27. Let f be as given in Theorem 3.1.26, then

$$f_{\text{SOS}} = f^* \iff \text{for all } \boldsymbol{\beta} \in Y,\ \boldsymbol{\beta} \in \Delta^* \text{ or } \boldsymbol{\beta} \in (2\mathbb{Z})^n \text{ and } b_{\boldsymbol{\beta}} > 0$$

Furthermore, if there exists $\boldsymbol{v} \in (\mathbb{R}^*)^n$ such that $b_{\boldsymbol{\beta}} < 0$ for all $\boldsymbol{\beta} \in Y$ then

$$f_{\text{SOS}} = f_{\text{SONC}} = f^*$$

Proof. Note that $f_{\text{SOS}} = f^*$ if and only if $f - f^* \in \Sigma_{n,2d}$. Since $f^* \in \mathbb{R}$, subtracting it from f does not effect the support of f. In addition, $f - f^*$ is SONC because f is SONC. Therefore Theorem 3.1.26 implies that $f - f^* \in \Sigma_{n,2d}$ if and only if $\boldsymbol{\beta} \in Y$, $\boldsymbol{\beta} \in \Delta^*$ or $\boldsymbol{\beta} \in (2\mathbb{Z})^n$ and $b_{\boldsymbol{\beta}} > 0$. Furthermore, if there exists $\boldsymbol{v} \in (\mathbb{R}^*)^n$ such that $b_{\boldsymbol{\beta}} < 0$ for all $\boldsymbol{\beta} \in Y$, then $f_{\text{SONC}} = f^*$ due to [IdW16b, Corollary 3.6]. □

Remark 3.1.28. We note that Corollary 3.1.27 generalizes [IdW16b, Corollary 3.6] since it shows that the converse implication also holds in the part (2) of [IdW16b, Corollary 3.6]. ⬡

3.2 MMS Preserving Functions and MMS Lattices

In this section we consider all simplicial basins $\Delta \subseteq \Delta_{2d}^n \cap \mathbb{Z}^n$ with $\mathbf{0} \in \Delta$, and give a classification of maximal mediated sets that arise from them. In order to do so, we first define a notion of density for MMS, see Definition 3.2.1. Then, we study the maps between simplicial basins that preserve our density notion. This yields an equivalence relation, and we point out that each equivalence class of simplicial basins can be identified with a lattice in $\mathbb{Z}^n$.

Let us start with fixing some extra notation that we require in this section. Given $\Delta = \{\boldsymbol{v(0)}, \ldots, \boldsymbol{v(n)}\} \subset (2\mathbb{N})^n$ a lexicographically ordered k-simplicial basin, then we

denote by M_Δ the column matrix of the elements in Δ, i.e.,

$$M_\Delta = \begin{bmatrix} \boldsymbol{v(0)} & \cdots & \boldsymbol{v(n)} \end{bmatrix}.$$

If $\boldsymbol{v(0)} = \mathbf{0}$ and $M_\Delta \in \mathbb{Z}^{n\times(n+1)}$, then we define the *lattice* associated to Δ as

$$L_\Delta := \langle \boldsymbol{r_1}, \ldots, \boldsymbol{r_n} \rangle \subset \mathbb{Z}^n, \tag{3.2.1}$$

where $\boldsymbol{r_i}$ is the i-th row of the matrix obtained by deleting the first column of M_Δ. Let $f \in P\left(\mathbb{R}[x_1, \ldots, x_n]_{2d}\right)$ be a circuit polynomial supported on a circuit with vertex set $\Delta = \operatorname{Vert}\left(\operatorname{New}\left(f\right)\right)$ and inner term $\boldsymbol{\beta} \in \mathbb{N}^n$. The maximal mediated set associated to f, Δ^*, is the set of choices for $\boldsymbol{\beta}$ that ensure $f \in \Sigma_{n,2d}$. Therefore, the density of Δ^* in $\operatorname{conv}(\Delta) \cap \mathbb{Z}^n$ is a measure of how likely f is to be a SOS polynomial. Due to Theorem 3.1.1, we have $\Delta \cup \operatorname{Mid}\left(\Delta\right) \subseteq \Delta^*$. Thus, we exclude these points that are a priori in the MMS while we describe the density of Δ^* in $\operatorname{conv}(\Delta) \cap \mathbb{Z}^n$.

Definition 3.2.1. Given a simplex $\Delta \subseteq (2\mathbb{Z})^n$, we define the *h-ratio* of Δ as follows:

$$\mathscr{H}\left(\Delta\right) = \begin{cases} \dfrac{\#\left(\Delta^* - \left(\Delta \cup \operatorname{Mid}\left(\Delta\right)\right)\right)}{\#\left(\left(\operatorname{conv}(\Delta) \cap \mathbb{Z}^n\right) - \left(\Delta \cup \operatorname{Mid}\left(\Delta\right)\right)\right)} & \text{if } \left(\operatorname{conv}(\Delta) \cap \mathbb{Z}^n\right) \neq \left(\Delta \cup \operatorname{Mid}\left(\Delta\right)\right), \\ 1 & \text{otherwise} \end{cases}$$

⬡

The h-ratio will be a significant statistical value for us, because the h-ratio of the Newton polytope is an indicator for the likelihood of a nonnegative circuit polynomial to be SOS.

Example 3.2.2. Let $\Delta_1 = \{(0,0), (0,2), (2,4)\}$ and $\Delta_2 = \{(0,0,0), (0,2,4), (0,6,4), (6,4,4)\}$. Then,

$$M_{\Delta_1} = \begin{bmatrix} 0 & 0 & 4 \\ 0 & 2 & 2 \end{bmatrix} \text{ and } M_{\Delta_2} = \begin{bmatrix} 0 & 0 & 0 & 6 \\ 0 & 2 & 6 & 4 \\ 0 & 4 & 4 & 4 \end{bmatrix}.$$

The lattices associated to Δ_1 and Δ_2 are

$$L_{\Delta_1} = \langle (0,4), (2,2) \rangle \text{ and } L_{\Delta_2} = \langle (0,0,6), (2,6,4), (4,4,4) \rangle.$$

The h-ratios corresponding to Δ_1 and Δ_2 are $\mathscr{H}\left(\Delta\right) = 1$ and $\mathscr{H}\left(\Delta\right) =$ We provide a visualization of Δ_1 along with M_{Δ_1} and L_{Δ_1} in Figure 3.5. ⬡

We aim to classify n-simplicial basins with maximal degree $2d$ according to their h-ratio. This classification will not only help us understand what properties of a simplicial

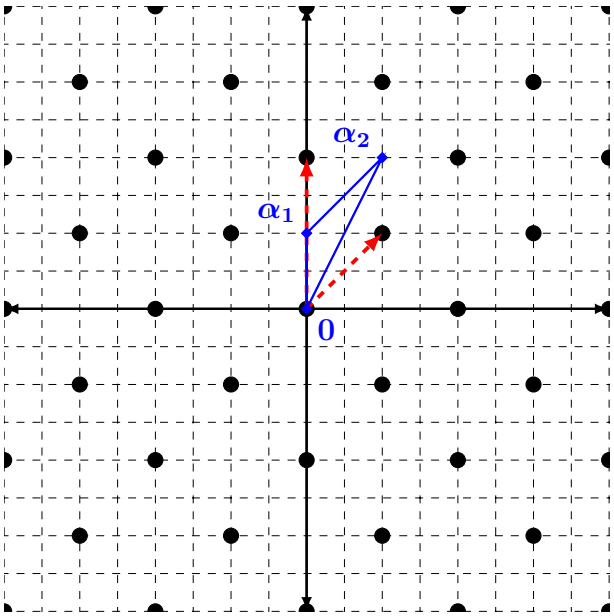

Figure 3.5: The simplicial basin Δ_1 from Example 3.2.2 is denoted as the blue triangle, the generators of L_{Δ_1} is denoted as red dashed vectors, and black dots denote the lattice L_{Δ_1}.

basin determine the h-ratio, but also it will yield an opportunity to reduce the size of the database of maximal mediated sets by storing one representative of the each class only. Therefore, in this section we study the maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ that preserve the maximal mediated set structure. In [Rez89, Page 445] the author points out that the maps that respect the MMS structure are necessarily linear maps in the context of trellises. We provide a rigorous proof for this observation in the setting of simplicial basins and h-ratios. In particular, we are interested in the following maps.

Definition 3.2.3. A function $T : \mathbb{R}^n \to \mathbb{R}^n$ is called *maximal mediated set preserving (MMS preserving)* if and only if it satisfies the following properties for every k-dimensional simplicial basin $\Delta \subseteq (2\mathbb{Z})^n$:

1. $T(\Delta) \subseteq (2\mathbb{Z})^n$ is a k-dimensional simplicial basin in $\mathbb{R}^n$.
2. For every $\boldsymbol{q} \in T(\Delta)^*$, there exists a unique $\boldsymbol{p} \in \Delta^*$ such that $T(\boldsymbol{p}) = \boldsymbol{q}$.
3. For every $\boldsymbol{q} \in (\mathrm{conv}(T(\Delta)) \cap \mathbb{Z}^n)$, there exists a unique $\boldsymbol{p} \in (\mathrm{conv}(\Delta) \cap \mathbb{Z}^n)$ such that $T(\boldsymbol{p}) = \boldsymbol{q}$.

⎔

Definition 3.2.3 has some immediate implications for every MMS preserving function T. Due to the first property with $k = 0$ we have:

$$\boldsymbol{p} \in (2\mathbb{Z})^n \implies T(\boldsymbol{p}) \in (2\mathbb{Z})^n. \tag{3.2.2}$$

The second and the third property respectively ensure for every k-dimensional simplicial basin Δ that

$$\#\Delta^* = \#T(\Delta)^* \text{ and } \#(\text{conv}(\Delta) \cap \mathbb{Z}^n) = \#(\text{conv}(T(\Delta)) \cap \mathbb{Z}^n).$$

Hence, the h-ratio is invariant under a maximal mediated set preserving function T.

Note that the property (1) of Definition 3.2.3 is equivalent to T mapping any k-dimensional affine independent subset of $(2\mathbb{Z})^n$ to a k-dimensional affine independent set of $(2\mathbb{Z})^n$. This means the restriction of T to $(2\mathbb{Z})^n$ is an affine transformation of $(2\mathbb{Z})^n$. The next proposition generalizes this result to $\mathbb{R}^n$.

Proposition 3.2.4. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a maximal mediated set preserving function, then T is a unimodular affine transformation. More specifically, we have

$$T(\boldsymbol{x}) = A_T\boldsymbol{x} + \boldsymbol{b}_T \tag{3.2.3}$$

with $\boldsymbol{b}_T \in (2\mathbb{Z})^n$, $A_T \in \mathbb{Z}^{n\times n}$, and $\det(A) = \pm 1$.

Proof. Assume that T is MMS preserving and let $K \subset \mathbb{R}^n$ be a collection of affine independent vectors. First, we show that $T(K)$ is also affine independent. This is true by Definition 3.2.3 for $K \subset (2\mathbb{Z})^n$ and thus also for $K \subset \mathbb{Q}^n$ with a suitable scaling of the elements of K. Since $\mathbb{Q}^n$ is a dense subset of $\mathbb{R}^n$, we conclude that T is an affine transformation over $\mathbb{R}^n$.

Due to (3.2.2) we know $T\left((2\mathbb{Z})^n\right) = (2\mathbb{Z})^n$. In particular, we have,

$$T(\mathbf{0}) = \boldsymbol{b}_T \in (2\mathbb{Z})^n.$$

Now we prove that $T(\mathbb{Z}^n) = \mathbb{Z}^n$, i.e., $A_T \in \mathbb{Z}^{n\times n}$. Let $\boldsymbol{p} = (p_1, \ldots, p_n) \in \mathbb{Z}^n$, and $J \subset [n]$ be the subset of indices where p_i is odd. We define two points $\boldsymbol{p}^+$ and $\boldsymbol{p}^-$ as follows:

$$p_i^+ = \begin{cases} p_i, \text{ if } i \notin J \\ p_i + 1, \text{ if } i \in J \end{cases}, \text{ and } p_i^- = \begin{cases} p_i, \text{ if } i \notin J \\ p_i - 1, \text{ if } i \in J \end{cases}$$

for all $i \in [n]$. Observe that $\boldsymbol{p}^+, \boldsymbol{p}^- \in (2\mathbb{Z})^n$ and $\boldsymbol{p} = \dfrac{\boldsymbol{p}^+ + \boldsymbol{p}^-}{2}$. Since T is affine we have,

$$T(\boldsymbol{p}) \;=\; T\left(\frac{\boldsymbol{p}^+ + \boldsymbol{p}^-}{2}\right) \;=\; \frac{T(\boldsymbol{p}^+) + T(\boldsymbol{p}^-)}{2}.$$

Due to (3.2.2), we have $T(\boldsymbol{p}^+), T(\boldsymbol{p}^-) \in (2\mathbb{Z})^n$, and thus $T(\boldsymbol{p}) \in \mathbb{Z}^n$. Therefore, $T(\mathbb{Z}^n) = \mathbb{Z}^n$ and hence $A_T \in \mathbb{Z}^{n\times n}$.

Finally, we show $\det(A_T) = \pm 1$. By part (1) of Definition 3.2.3, T maps any set of n linearly independent vectors to another set of n linearly independent vectors. Thus, $\det(A_T) \neq 0$. By part (3) of Definition 3.2.3 volumes of simplices are preserved under T, and hence $\det(A_T) = \pm 1$. □

Recall that any affine linear transformation can be represented as an element of the group $\mathbb{R}^n \rtimes \mathrm{GL}(\mathbb{R}^n)$. In particular, if T is MMS preserving, then we have $T \in (2\mathbb{Z})^n \rtimes \mathrm{GL}(\mathbb{Z}^n)$ due to Proposition 3.2.4. This motivates the following definition.

Definition 3.2.5. We define the *maximum mediated set preserving group* of $\mathbb{R}^n$, $\mathscr{M}_{\mathrm{n}}$ as follows:

$$\mathscr{M}_{\mathrm{n}} \;=\; \left(\{T \in (2\mathbb{Z})^n \rtimes \mathrm{GL}(\mathbb{Z}^n) \;\mid\; T \text{ is maximal mediated set preserving } \}\,, \circ\right),$$

where $\circ$ denotes the usual composition of functions. ⬡

Obviously, $\mathscr{M}_{\mathrm{n}}$ is a subgroup of $(2\mathbb{Z})^n \rtimes \mathrm{GL}(\mathbb{Z}^n)$. In the next theorem we show that it is in fact the full group.

Theorem 3.2.6. $T \in \mathscr{M}_{\mathrm{n}}$ if and only if $T \in (2\mathbb{Z})^n \rtimes \mathrm{GL}(\mathbb{Z}^n)$, i.e., $T : \mathbb{R}^n \to \mathbb{R}^n$ is MMS preserving if and only if

$$T(\boldsymbol{x}) \;=\; A_T\boldsymbol{x} + \boldsymbol{b}_T$$

is a unimodular affine transformation with $\boldsymbol{b}_T \in (2\mathbb{Z})^n$.

Proof. Let $T \in \mathscr{M}_{\mathrm{n}}$, the only if part of the theorem follows from Proposition 3.2.4. For the converse, assume that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a unimodular affine transformation with $\boldsymbol{b}_T \in (2\mathbb{Z})^n$. We need to show that T is MMS preserving. Let $\Delta = \{\boldsymbol{v_0}, \ldots, \boldsymbol{v_k}\}$ be a k-simplicial basin. By definition, $\boldsymbol{v_0}, \ldots, \boldsymbol{v_k}$ are affine independent, and, since T is affine, $T(\Delta)$ is a set of $k+1$ affine independent vectors, i.e., a k-simplicial basin. Furthermore, if $\Delta \subset (2\mathbb{Z})^n$, then $T(\Delta) \subset (2\mathbb{Z})^n$ because

$$T(\boldsymbol{v_i}) \;=\; A_T\boldsymbol{v_i} + \boldsymbol{b}_T \in (2\mathbb{Z})^n.$$

This implies part (1) of Definition 3.2.3.

Next we show part (3) of Definition 3.2.3: Let $\boldsymbol{q} \in \operatorname{conv}(T(\Delta)) \cap \mathbb{Z}^n$. Since T is unimodular, T^{-1} exists and is also unimodular. This implies that $T^{-1}(\boldsymbol{q}) \in \mathbb{Z}^n$. Furthermore since $\boldsymbol{q} \in \operatorname{conv}(T(\boldsymbol{v_0}), \ldots, T(\boldsymbol{v_n}))$, there exists $\lambda_0, \ldots, \lambda_n \in \mathbb{R}_{>0}$ with $\sum_{i=0}^n \lambda_i = 1$ and $\boldsymbol{q} = \sum_{i=0}^n \lambda_i T(\boldsymbol{v_i})$. Then,

$$T^{-1}(\boldsymbol{q}) \;=\; T^{-1}\left(\sum_{i=0}^{n} \lambda_i T(\boldsymbol{v_i})\right) \;=\; \sum_{i=0}^{n} \lambda_i T^{-1}(T(\boldsymbol{v_i})) \;=\; \sum_{i=0}^{n} \lambda_i \boldsymbol{v_i} \in \operatorname{conv}(\boldsymbol{v_0}, \ldots, \boldsymbol{v_k}).$$

Therefore, for all $\boldsymbol{q} \in \operatorname{conv}(T(\Delta)) \cap \mathbb{Z}^n$, there exists a unique $\boldsymbol{p} = T^{-1}(\boldsymbol{q}) \in \operatorname{conv}(\Delta) \cap \mathbb{Z}^n$.

Finally, we show part (2) of Definition 3.2.3: Since T is a bijective map between $\operatorname{conv}(\Delta) \cap \mathbb{Z}^n$ and $\operatorname{conv}(T(\Delta)) \cap \mathbb{Z}^n$, we are done if we show that $\boldsymbol{p} \in \Delta^*$ if and only if $T(\boldsymbol{p}) \in T(\Delta)^*$. Define the sets $U^0 = \operatorname{conv}(\Delta) \cap \mathbb{Z}^n$, $V^0 = \operatorname{conv}(T(\Delta)) \cap \mathbb{Z}^n$, and define the sets U^k and V^k recursively as follows:

$$U^k \;=\; \operatorname{Mid}\left(U^{k-1}\right) \cup \Delta, \qquad V^k \;=\; \operatorname{Mid}\left(V^{k-1}\right) \cup T(\Delta).$$

By Algorithm 3.1.12, we know that $\boldsymbol{p} \in \Delta^*$ if and only if $\boldsymbol{p} \in U^k$ for all k and $T(\boldsymbol{p}) \in T(\Delta)^*$ if and only if $T(\boldsymbol{p}) \in V^k$ for all k. We claim that $T(U^k) = V^k$ is a bijection for all $k \in \mathbb{N}$ and we argue by induction over k. We already know that $T(U^0) = V^0$ is a bijection. Now assume that T sends U^k to V^k bijectively, and let $\boldsymbol{q}$ be a point in V^{k+1}. Then either $\boldsymbol{q} \in \operatorname{Mid}\left(V^k\right)$ or $\boldsymbol{q} \in T(\Delta)$. On the one hand, if $\boldsymbol{q} \in T(\Delta)$, then $\boldsymbol{q}$ is a vertex of $\operatorname{conv}(T(\Delta))$ and hence there exists a unique $\boldsymbol{p} \in \Delta$ with $T(\boldsymbol{p}) = \boldsymbol{q}$ by part (1) of Definition 3.2.3, which we have already shown to hold.

On the other hand, if $\boldsymbol{q} \in \operatorname{Mid}\left(V^k\right)$, then there exist distinct $\widetilde{\boldsymbol{s}}, \widetilde{\boldsymbol{t}} \in V^k$ such that

$$\boldsymbol{q} \;=\; \frac{1}{2}(\widetilde{\boldsymbol{s}} + \widetilde{\boldsymbol{t}}).$$

By the induction hypothesis, $\widetilde{\boldsymbol{s}}$ and $\widetilde{\boldsymbol{t}}$ have unique preimages $\boldsymbol{s}$ and $\boldsymbol{t}$ in U^k respectively. Since T^{-1} is affine linear, we obtain a unique

$$\boldsymbol{p} \;=\; T^{-1}(\boldsymbol{q}) \;=\; T^{-1}\left(\frac{1}{2}(\widetilde{\boldsymbol{s}} + \widetilde{\boldsymbol{t}})\right) \;=\; \frac{1}{2}\left(T^{-1}(\widetilde{\boldsymbol{s}}) + T^{-1}(\widetilde{\boldsymbol{t}})\right) \;=\; \frac{1}{2}(\boldsymbol{s} + \boldsymbol{t}).$$

Thus, for all k, T maps U^k to V^k bijectively. Hence, we conclude $\boldsymbol{p} \in \Delta^*$ if and only if $T(\boldsymbol{p}) \in T(\Delta)^*$. □

We present a key corollary of Theorem 3.2.6. If we exclude the translations, e.g. by considering only those simplicial basins that contain $\mathbf{0}$, then an MMS preserving function is given by a unimodular matrix A_T. We show that the row span of $A_T M_\Delta$ yields,

up to a permutation of the coordinates, the same lattice L_Δ as the row span of M_Δ. Therefore, $\mathscr{H}(\Delta)$ of a k-simplicial basin Δ containing $\mathbf{0}$ is actually an invariant of the lattice generated by the rows of M_Δ. Before we state this result with more rigorous terms in Corollary 3.2.7, we first recall the notion of Hermite normal form. Given a full column rank matrix $M \in \mathbb{Z}^{m\times n}$, its *Hermite normal form* is given by a full column rank matrix $H \in \mathbb{Z}^{m\times n}$ along with the unimodular companion matrix $U \in \mathbb{Z}^{n\times n}$ such that $M = UA$ and,

1. H is upper triangular matrix such that all zero rows of H appear below nonzero rows,

2. The pivot element of each nonzero row , i.e. the first element from left that is not zero, is positive,

3. The pivot element of each nonzero row is located in the main diagonal of H, and it is the maximal element of its column.

We note two fact without proof: First, every full column rank $M \in \mathbb{Z}^{m\times n}$ has a unique Hermite normal form, and second, the rows $M_1, M_2 \in \mathbb{Z}^{m\times n}$ generate the same lattice if and only if their Hermite normal form are same. For more information on Hermite normal form, and for the proofs of these two facts, we refer to [Sch11, Section 4.1]

Corollary 3.2.7. Let $\Delta_1 = \{\mathbf{0}, \boldsymbol{v_1}, \dots, \boldsymbol{v_k}\}$ and $\Delta_2 = \{\mathbf{0}, \boldsymbol{u_1}, \dots, \boldsymbol{u_k}\}$ be two k-simplicial basins in $2\mathbb{Z}^n$. Then, there exists a $T \in \mathscr{M}_{\mathrm{n}}$ with $\boldsymbol{b}_T = \mathbf{0}$ such that

$$T(\Delta_1)^* \;=\; \Delta_2^*$$

if and only if the lattices L_{Δ_1} and L_{Δ_2} share the same Hermite normal form up to a permutation of columns.

Proof. If there exists $T \in \mathscr{M}_{\mathrm{n}}$ with $\boldsymbol{b}_T = \mathbf{0}$ and $T(\Delta_1)^* = \Delta_2^*$, then $M_{\Delta_1} = A_T M_{\Delta_2}$ where $A_T \in \mathrm{GL}(\mathbb{Z}^n)$. Therefore, the $\mathbb{Z}$-row span of M_{Δ_1} and M_{Δ_2} yields, up to a permutation of the coordinates, the same lattice. In converse, if the lattices L_{Δ_1} and L_{Δ_2} share the same Hermite Normal Form up to a permutation of columns, then the $\mathbb{Z}$-row spans of M_{Δ_1} and M_{Δ_2} coincide up to a permutation of columns. Hence, there exists $A \in \mathrm{GL}(\mathbb{Z}^n)$ such that $M_{\Delta_1} = AM_{\Delta_2}$. Thus, the transformation $T(\boldsymbol{x}) = A\boldsymbol{x}$ is MMS preserving. □

Corollary 3.2.7 will be quite practical for reducing the size of the database in Section 3.3, see e.g. Table 3.5. To conclude this section, we give two examples to illustrate the lattices mentioned in Corollary 3.2.7.

Example 3.2.8. Let

$$\mathbf{0} = \begin{bmatrix}0\\0\end{bmatrix}, \boldsymbol{v_1} = \begin{bmatrix}2\\4\end{bmatrix}, \boldsymbol{v_2} = \begin{bmatrix}4\\2\end{bmatrix}, \boldsymbol{u_1} = \begin{bmatrix}2\\0\end{bmatrix}, \boldsymbol{u_2} = \begin{bmatrix}4\\6\end{bmatrix},$$

and $\Delta_1, \Delta_2 \subset (2\mathbb{Z})^2$ be given as $\Delta_1 = \{\mathbf{0}, \boldsymbol{v_1}, \boldsymbol{v_2}\}$ and $\Delta_2 = \{\mathbf{0}, \boldsymbol{u_1}, \boldsymbol{u_2}\}$. Then we write the matrices M_{Δ_1} and M_{Δ_2} with the given order;

$$M_{\Delta_1} = \begin{bmatrix}2 & 4\\4 & 2\end{bmatrix} \quad \text{and} \quad M_{\Delta_2} = \begin{bmatrix}2 & 4\\0 & 6\end{bmatrix}.$$

If we denote the Hermite normal forms of M_{Δ_1} and M_{Δ_2} as H_1 and H_2 respectively, then we have

$$H_1 = \begin{bmatrix}2 & 4\\0 & 6\end{bmatrix} = \begin{bmatrix}1 & 0\\2 & -1\end{bmatrix} M_{\Delta_1} \quad \text{and} \quad H_2 = \begin{bmatrix}2 & 4\\0 & 6\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} M_{\Delta_2}.$$

Therefore, the lattices corresponding to the row spans of M_{Δ_1} and M_{Δ_1} share the same Hermite normal form, and they are the same lattice, i.e.,

$$L_{\Delta_1} = \left\langle \begin{bmatrix}2\\4\end{bmatrix}, \begin{bmatrix}4\\2\end{bmatrix} \right\rangle = \left\langle \begin{bmatrix}2\\4\end{bmatrix}, \begin{bmatrix}0\\6\end{bmatrix} \right\rangle = L_{\Delta_2}.$$

Furthermore, if we consider map $T(\boldsymbol{x}) = Ax$ defined by the unimodular matrix

$$A = \begin{bmatrix}1 & 0\\2 & -1\end{bmatrix},$$

then we see that $M_{\Delta_2} = AM_{\Delta_1}$, and $T(\Delta_1)^* = \Delta_2^*$. See also Example 3.2.8 for a visualization of this example. ⎔

Note that in the Example 3.2.8 we obtain identical lattices. We show that this is not always the case.

Example 3.2.9. Let

$$\mathbf{0} = \begin{bmatrix}0\\0\end{bmatrix}, \boldsymbol{v_1} = \boldsymbol{u_2} = \begin{bmatrix}2\\2\end{bmatrix}, \boldsymbol{v_2} = \boldsymbol{u_1} = \begin{bmatrix}0\\6\end{bmatrix},$$

and $\Delta_1, \Delta_2 \subset (2\mathbb{Z})^2$ be given as $\Delta_1 = \{\mathbf{0}, \boldsymbol{v_1}, \boldsymbol{v_2}\}$ and $\Delta_2 = \{\mathbf{0}, \boldsymbol{u_1}, \boldsymbol{u_2}\}$. Δ_1 and Δ_2 have the same maximal mediated sets since they correspond to the same simplex with different

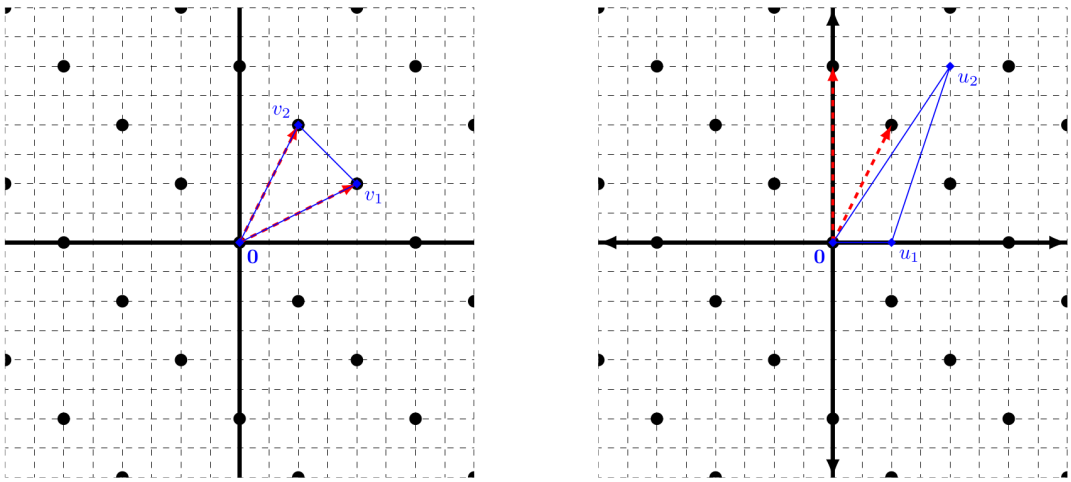

Figure 3.6: Black points correspond to the lattices generated by the rows of L_{Δ_1} (left) and L_{Δ_2} (right) where Δ_1 and Δ_2 are given as in Example 3.2.8. The (blue) triangles visualizes the 2-simplicial basins Δ_1, Δ_2 and the (red) dashed vectors shows the generators of L_{Δ_1} and L_{Δ_2}.

vertex order. Again, we write the matrices,

$$M_{\Delta_1} = \begin{bmatrix} 2 & 0 \\ 2 & 6 \end{bmatrix} \quad \text{and} \quad M_{\Delta_2} = \begin{bmatrix} 0 & 2 \\ 6 & 2 \end{bmatrix}.$$

If we denote the Hermite Normal Form of M_{Δ_1} and M_{Δ_2} as H_1 and H_2 respectively, then we have

$$H_1 = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} M_{\Delta_1} \quad \text{and} \quad H_2 = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} M_{\Delta_2}.$$

In this case Hermite Normal Forms are equal up to permutation of columns. Therefore, the lattices generated by the row spans of M_{Δ_1} and M_{Δ_2} are not identical, but they are isomorphic. This isomorphism is given by a permutation of the coordinates of the lattice, see Figure 3.7.

⬡

3.3 Maximal Mediated Set Database

One of the major goals of [HRdWY20] is to generate a database by classifying all maximal mediated sets of n-simplicial basins with maximal degree $2d$ for n and $2d$ as large as possible. In this section, we first describe the methodology of our approach to

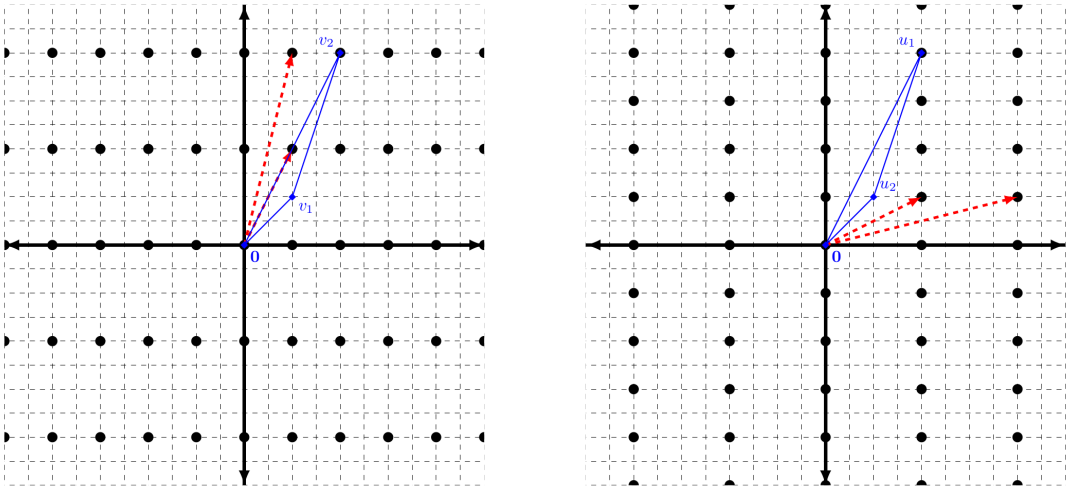

Figure 3.7: Black points correspond the lattices L_{Δ_1} (left) and L_{Δ_2} (right) where Δ_1 and Δ_2 are given as in Example 3.2.9. The (blue) triangles visualizes the 2-simplicial basins Δ_1, Δ_2 and the (red) dashed vectors visualizes the generators of L_{Δ_1} and L_{Δ_2}.

generate the database. There are three major parts to the process:

1. Enumerating simplices (Section 3.3.1),
2. classifying simplices (Section 3.3.2),
3. computing the maximal mediated set (Section 3.3.3).

Then in Section 3.3.4, we provide the experimental setup that we used while generating the database. As conclusion, we give an analysis and some significant statistics of the database in Section 3.3.5.

We note that the implementation of related algorithms to form the database was done in `C++` using the Polymake [GJ00] software package; its source code can be obtained via

https://polymake.org/doku.php/extensions/max_mediated_sets

The database itself, and the instructions manual are available in the following link:

https://polymake.org/downloads/MMS/

3.3.1 Enumerating Simplices

In order to explain the enumeration process, we introduce some notation exclusive to Section 3.3 Assume that n and $2d$ are fixed. Following Corollary 3.2.7, we restrict to the

simplices containing a vertex at the origin. Let

$$V \;=\; \left[(x_1,\ldots,x_n)\in(2\mathbb{N})^n \mid \sum_{i=1}^{n} x_i \le 2d\right] - \mathbf{0}$$

denote the lexicographically ordered list of all even lattice points (excluding $\mathbf{0}$) with 1-norm at most $2d$. In what follows, we represent V as a $(\binom{n+2d-1}{2d}-1)\times n$ matrix where the j-th row contains the j-th entry in the list. Generating all n-simplices containing the origin thus is equivalent to listing all full-rank $n\times n$ submatrices of V. In order to reduce the number of necessary rank computations, we construct submatrices row by row, adding rows in the order of V. Given a *k-index set* $J\subset[\#V]$, $\#J=k$, we denote its entries by $J_1,\ldots,J_k$ and we denote by V_J the submatrix of V formed by the row indices J. We call a k-index set J a *prefix* of an n-index set M if $k\le n$ and $J_i=M_i$ for $i\in[k]$. Before we move on to the database, we introduce some additional notation exclusive to Section 3.3.1.

If we have an n-index set I, then the following algorithm computes the lexicographically next n-index set J such that $\operatorname{rank}(V_J)=n$.

Algorithm 3.3.1. **Input:** V: Matrix of valid simplex vertices, I: n-index set
Output: J: the lex-next n-index set with $\operatorname{rank}(V_J)=n$, if such a set exists; $\emptyset$: otherwise

1: $J \leftarrow$ lexicographic successor of I
2: $K \leftarrow \emptyset$
3: **for** $j\in[n]$ **do**
4: $\quad K \leftarrow K\cup\{J_j\}$
5: $\quad$ **if** $\operatorname{rank}(V_K)<\#K$ **then**
6: $\quad\quad$ **if** there exists a $\#K$-index set $\widehat{K}$ on $[\#V-\#K]$ with $\widehat{K}>_{\text{lex}}K$ **then**
7: $\quad\quad\quad J\leftarrow \min_{lex}\left\{\widehat{K}\cup M \text{ } n\text{-index set} : \widehat{K}>_{\text{lex}}K, \max(\widehat{K})<\min(M)\right\}$
8: $\quad\quad\quad K\leftarrow$ largest prefix of J contained in K
9: $\quad\quad\quad j\leftarrow \#K+1$
10: $\quad\quad$ **else**
11: $\quad\quad\quad$ **return** $\emptyset$
12: $\quad\quad$ **end if**
13: $\quad$ **end if**
14: **end for**
15: **return** J

Proof. The correctness of the algorithm is clear by construction. In line 7, J is lexicographically increased, and K is always a prefix of J. The condition in line 6 does not hold for any prefix of the lexicographically maximal n-index set on $[\#V]$, hence the algorithm

terminates. □

We note that Algorithm 3.3.1 was constructed, and implemented by Olivia Röhrig, see [Roe20, Algorithm 2.1.1].

Let K be an index set and $l \in [\#V]$ such that $l \notin K$. If $\operatorname{rank}(V_K) = \operatorname{rank}(V_{K\cup\{l\}})$, then Algorithm 3.3.1 excludes all instances containing $K \cup \{l\}$. Therefore, we avoid the rank checks for all further matrices containing the $V_{K\cup\{l\}}$. For the enumeration process, we use Algorithm 3.3.1 in parallel by assigning to one thread the enumeration of all matrices that have a distinguished $p \in V$ as the their first row. The threads can then be run as independent processes as no inter-thread communication is required. We did so using the `GNU parallel` software [Tan11].

3.3.2 Classifying Simplices

Two different simplicial basins $\Delta_1, \Delta_2 \subseteq (2\mathbb{Z})^n$ may have the same maximal mediated set structure, i.e., there exists a $T \in \mathscr{M}_n$ such that $T(\Delta_1) = \Delta_2$. Corollary 3.2.7 implies that h-ratio is an invariant of an underlying lattice rather than of the simplicial basin itself. Therefore, instead of the distribution of h-ratios over n-simplicial basins with maximal degree $2d$, we consider the distribution of h-ratios over possible lattices that can arise from n-simplicial basins with maximal degree $2d$. For any n-simplicial basin Δ, we want to find a unique representative from the set $\{T(\Delta) | T \in \mathscr{M}_n\}$. Unfortunately, computing the Hermite normal form straightforwardly is not sufficient to find a unique representation, because as shown in Example 3.2.9, reordering the vertices yields different Hermite normal forms. Hence, we consider all orderings, compute their Hermite normal forms, and check whether we have already encountered that lattice before.

Remark 3.3.2. Let S be a finite set of n-simplicial basins. The *Hermite normal form reduction (HNF reduction)* of S is the process of identifying every n-simplicial basin in S if they reduce to the same (row) Hermite normal form up to permutation of columns. It is straightforward to see that this process divides S into equivalence classes. ⬡

The aforementioned HNF reduction in Remark 3.3.2 is only feasible if a fast lookup of previously encountered lattices is available. As for the instances of interest the set of these is larger than the memory available to us, we had to resort to an on-disk key-value store. We also required support for deadlock-free lookups and writes from multiple threads so we could still benefit from the parallelizability of the enumeration. We ended up using a `BerkeleyDB` [OBS99] database, storing Hermite normal form as key and their maximal mediated set, companion matrices, h-ratio and other interesting information as value.

3.3.3 Computing MMS

After enumerating and dividing the set of all n-simplicial basins of maximal degree $2d$ into equivalence classes, we simply choose the lex-minimal element in each class as representative. In Section 3.1, we introduced two algorithms Algorithm 3.1.12 and Algorithm 3.1.14, and showed that both algorithms compute the MMS of a given simplicial basin. We use our `POLYMAKE` extension (see Remark 3.3.3), which incorporates Algorithm 3.1.14, to compute the maximal mediated set of simplicial basins.

Remark 3.3.3. We note that, Algorithm 3.1.14 is implemented in Polymake as an extension, which is available in the following link:

https://polymake.org/doku.php/extensions/max_mediated_sets

⎔

There are a two additional aspects of our implementation algorithm that we did not highlight in the pseudo code in Algorithm 3.1.14.

(1) To increase efficiency, we incorporated cost efficient pre-computation checks based on [Rez89, Theorem 2.5, Theorem 2.7]. They let us detect H-simplices and M-simplices without going through the iteration process in some cases.

(2) We keep L in lexicographical order, which ensures that for any point $i \in L$, if it is midpoint of two points in L then one of those two will appear before i in the list and will appear after i. This enables us to find out whether $i \in \mathrm{Mid}(L)$ in at most $\frac{1}{4}(\#L)^2$ operations.

3.3.4 Experimental Setup

Here we give the overview of the experimental setup we have used during generation of the database.

Software We performed the maximal mediated set computations in the open source software Polymake. We have written our an extension to Polymake that computes the maximal mediated set using Algorithm 3.1.14, see Remark 3.3.3.

Investigated Data The investigated data consists of simplicial basins and the lattices underlying simplicial basins described in Corollary 3.2.7. We divide the investigated data into smaller sets according to two parameters:

n the dimension of the simplicial basins.

dimension	2	3	4	5	7
degree	150	16	14	8	4

Table 3.1: Maximal degrees and dimensions of the fully computed data.

dimension	4	5	6	7	8	9
degree	16	16	16	16	16	16

Table 3.2: Maximal degrees and dimensions of the sampled data.

2d the maximal total degree of the simplicial basins, i.e. the minimum integer for a given simplicial basin Δ such that for all $\boldsymbol{p} \in \operatorname{conv}(\Delta)$, $\mathbf{1}^T \cdot \boldsymbol{p} < 2d$.

Table 3.1 summarizes for which $n, 2d \in \mathbb{N}$ we have computed MMS of every possible instance. We resort to sampling instead of computing the maximal mediated set for all simplicial basins for larger n and $2d$. Table 3.2 summarizes the cases for which values of $n, 2d \in \mathbb{N}$ we have sampled n-simplicial basins $\Delta \subset (2\mathbb{Z})^n$ such that zero vector is a vertex of $\operatorname{conv}(\Delta)$.

Remark 3.3.4. We stored each MMS that we compute, and constructed a database for the cases given in Table 3.1 and Table 3.2. The database is available at

https://polymake.org/downloads/MMS/

⬡

Sampling For reproducibility of our sampling, we provide the `RandomSimplexIterator` class in our Polymake package, using the type `UniformlyRandom<Integer>` in Polymake. This class initially produces an array of integers from a seeded uniform distribution with the seed 1. Then it picks n points in $\mathbb{N}^n$ with 1-norm at most $2d$ uniformly at random seeded by the elements the integer array. If these n points together with the origin are affine independent then we keep this sample, otherwise we discard it and pick another n point uniformly at random. We did not set a specific sample size as stopping criterion, since we aim to compute as many MMSes as we can. Thus, we stopped the sampling process according to elapsed time for each n and $2d$, see the point Sampled Data Sets in Section 3.3.5.

Hardware and System We used three separate computers for the computations. For $n = 4$ and $2d = 14$, we used a `AMD Phenom(TM) II X6 1090T` with 5 cores, 16 GB of RAM under openSUSE Leap 15.0. For $n = 5$ and $2d = 8$, we used a `AMD Phenom(TM) II X6 1090T` with 6 cores, 16 GB of RAM under openSUSE Leap 42.3. For the remaining computations, we used `Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz` with 12 cores, 16 GB of RAM under openSUSE Leap 15.0.

3.3.5 Computational Results and Database Statistics

In this section, we analyze the experimental results we achieved. Our main measurement is the h-ratio defined in Definition 3.2.1, and we are interested how the h-ratio is distributed for fixed dimension and degree.

Maximal Mediated Subsets of 2-Simplicial Sets First we address Conjecture 3.1.19, which was announced in [Rez89, Page 9]. We computed the maximal mediated sets of all 4266834 2-simplicial basins with maximal degree 150, and confirmed that Conjecture 3.1.19 holds. These 2-simplicial basins arise from 886297 different lattices as described in Corollary 3.2.7 after an Hermite normal form reduction. We summarize the statistics corresponding to $n = 2$, $2d = 150$ case in Table 3.3. From the

	Total count	H-simplex	M-simplex	mean of h-ratio	SD of h-ratio
2-Simplicial Sets	4266834	4250533	16301	0.996179	0.061691
Lattices	886297	886188	109	0.999877	0.011089

Table 3.3: From left to right, the total number of objects, the number of H-simplices, the number of M-simplices, the average h-ratio, and the standard deviation of the h-ratio for 2-simplicial basins and their underlying lattices in the case $n = 2$, $2d = 150$.

Table 3.3, we see that number of H-simplices is significantly higher than number of M-simplices; a fact suggested by [IdW16a, Theorem 5.9]. Thus, we provide experimental evidence that a clear majority of all nonnegative circuit polynomials in $\mathbb{R}[x, y]$ are sums of squares.

Fully Computed Data Sets for Higher Dimensions Here we focus on the fully computed cases, i.e,. the cases of n and $2d$ in the database where we were able to exhaust all simplicial basins. We point out that in these cases the Hermite normal form reduction indeed yields a different distribution for h-ratio. First, we summarize the general statistics we obtained from the fully computed data sets in Table 3.4.

We observe that the number of underlying lattices is significantly smaller than the number of simplicial basins. To point out this difference more rigorously, we provide in Table 3.5 the factors of decrease for each n and $2d$ in the fully computed cases.

Table 3.5 reveals that the decrease factor is more sensitive to a change in n than a change in $2d$. We explain this experimental observation theoretically in what follows. Let Δ be a n-simplicial basin and consider the coordinate permutation given by $\Pi_\sigma : x_i \mapsto x_{\sigma(i)}$ where $\sigma \in S_n$. Observe that:

1. Since $\Pi_\sigma \in \mathscr{M}_{\mathrm{n}}(\mathbb{R}^n)$, Δ and $\Pi_\sigma(\Delta)$ share the same Hermite normal form after a suitable permutation of columns. Therefore, Δ and $\Pi_\sigma(\Delta)$ are in the same class of lattices in our database.

n	2d		Total Count	Mean of h-ratio	SD of h-ratio
3	10	Simplicial Set	21636	0.724138	0.392967
		Lattice	782	0.592994	0.397988
3	16	Simplicial Set	659082	0.638828	0.412316
		Lattice	20429	0.583357	0.412889
4	14	Simplicial Set	853024289	0.433506	0.383378
		Lattice	1602368	0.227706	0.273419
5	8	Simplicial Set	305565979	0.680445	0.373089
		Lattice	53306	0.470493	0.303315
7	4	Simplicial Set	2414505	0.931788	0.238172
		Lattice	19	0.853923	0.304942

Table 3.4: This table summarizes the statistics of n-simplicial basins with maximal degree $2d$ and the statistics of lattices underlying n-simplicial basins with maximal degree $2d$ for fully computed data sets.

n	2d	Decrease Factor
3	10	27.667519
3	16	32.262078
4	14	532.352299
5	8	5732.299910
7	4	127079.210526

Table 3.5: The factors of decrease in the number of stored maximal mediated sets after a Hermite normal form reduction is performed.

2. Maximal degrees of Δ and $\Pi_\sigma(\Delta)$ are equal.
3. Increasing the dimension from n to $n+1$ yields $(n+1)! - n!$ new possible coordinate permutations.

Therefore, increasing n increases the number of simplicial basins with the same Hermite normal form exponentially even if the maximal degree $2d$ is fixed. However, increasing $2d$ while n is fixed does not create new symmetries and hence does not affect the decrease factor as severely. The decrease in the total count of objects in Table 3.4 is crucial to reduce the size of the database. However, the HNF reduction is computationally costly since making use of Corollary 3.2.7 requires to consider all column permutations of the underlying matrix. Thus, there is a trade off between the memory, which is needed for the database, and the time needed to compute the database.

Recall that MMS is an invariant of the underlying lattice of its defining simplex by Corollary 3.2.7. Therefore, considering lattices (instead of the original simplicial

basins) yields a more accurate description of the behavior of the h-ratio. We plot the distributions of simplicial basins and lattices in Figure 3.8 to visualize the significant difference between the distributions. The difference of the distributions follows

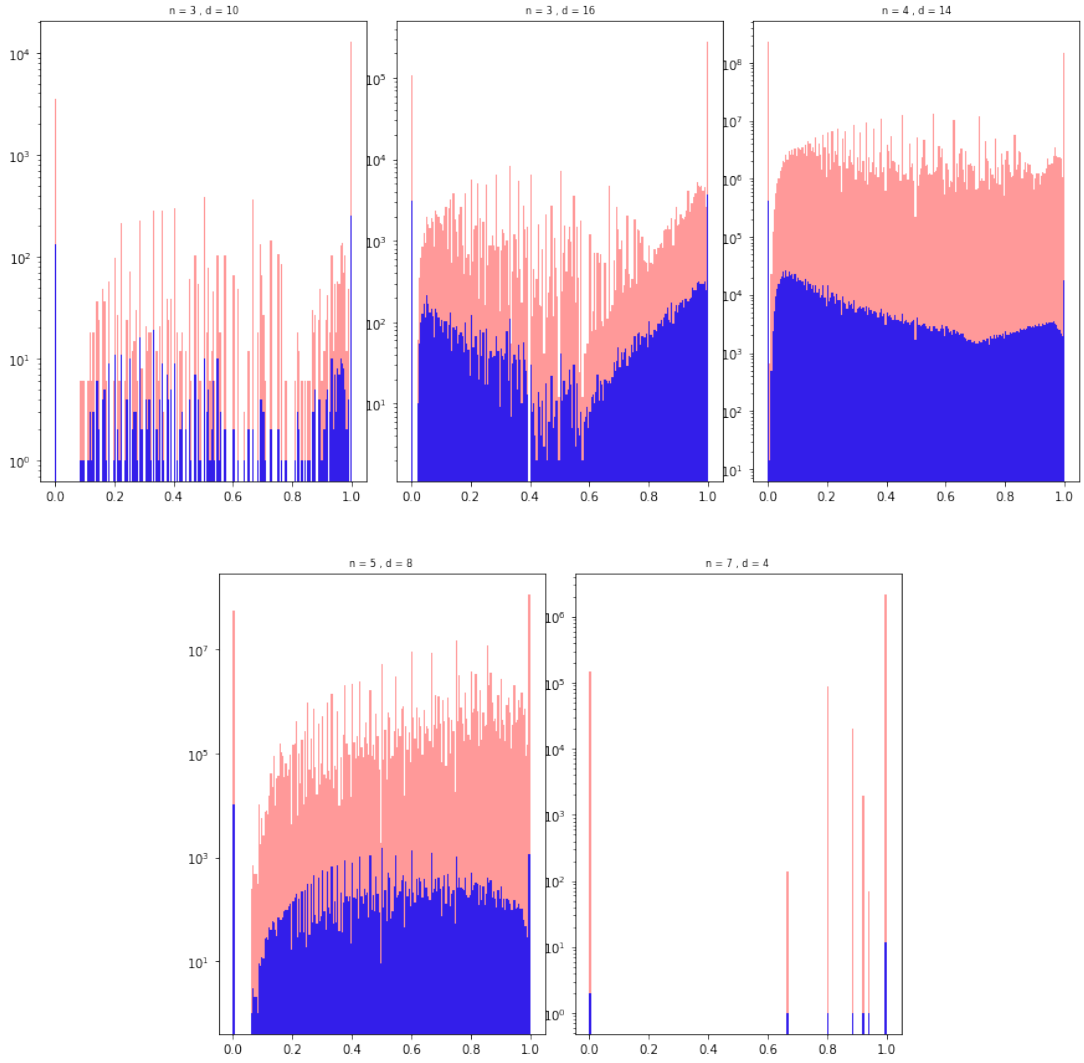

Figure 3.8: The distribution of h-ratio over the n-simplicial basins (red), and over lattices (blue) for the given n and $2d$ in Table 3.1.

moreover from our results in Table 3.4 and Table 3.5. More specifically, we observe that Hermite normal form reduction decreases the expected h-ratio. The standard deviation for simplicial basins is more stable compared to standard deviation for lattices.

Sampled Data Sets For higher n and $2d$ we produced a sampled database as described

n	2d		Total Count	Mean of h-ratio	SD of h-ratio
4	16	Simplicial Set	10000000	0.392896	0.370466
		Lattice	2067884	0.221803	0.277060
5	16	Simplicial Set	5000000	0.299490	0.320094
		Lattice	3297468	0.216518	0.245109
6	16	Simplicial Set	1000000	0.290170	0.322581
		Lattice	904317	0.263667	0.297612
7	16	Simplicial Set	100000	0.325715	0.361047
		Lattice	98016	0.319911	0.356507
8	16	Simplicial Set	10000	0.387188	0.411047
		Lattice	9966	0.385937	0.410454
9	16	Simplicial Set	1302	0.625456	0.447447
		Lattice	1000	0.512343	0.453334

Table 3.6: This table summarizes the statistics of n-simplicial basins with maximal degree $2d$ and the statistics of lattices underlying n-simplicial basins with maximal degree d for fully computed data sets.

in the sampling part of Section 3.3.4. In Table 3.6 we present the summary of h-ratio statistics of the sampled cases for simplicial basins and underlying lattices and in Figure 3.9 we plot the distributions of simplicial basins and underlying lattices.

We note that the sampled data for $n = 8$ and $n = 9$ are incomplete, and hence do not provide much information about the statistics. However we include them in Table 3.6 and Figure 3.9, to underline the fact that our sampling approach does not operate smoothly with HNF reduction for higher values of n. This is due to the exponential increase in the column permutations that one has to check for HNF reduction.

While the sampled data does not yield a clear, noise-free distribution, we still observe a similar trend in the expected h-ratios for $n \in \{4, 5, 6, 7\}$. Analogously to the fully computed (i.e., nonsampled) cases, the expected h-ratio of lattices are smaller than expected h-ratio of simplices also in the sampled cases.

Further Effects of HNF Reduction to Statistics We observe peaks in the graphs provided in Figure 3.8. In order to study these peaks more closely, we plot the h-ratio distributions of the case $n = 4$ as $2d$ increases in Figure 3.10. For small values of $2d$ we have individual peaks because a small maximal degree only allows a few different h-ratios to occur. For sufficiently large $2d$ the distribution becomes visible, since larger variety of h-ratios can appear. Furthermore, as $2d$ increases, the individual peaks that exists for smaller $2d$ survive and form the spikes we observe in the red distributions. Figure 3.10 illustrates how as $2d$ increases, individual peaks

transform into a noisy distribution with spikes.

Even though the spikes are present in distribution of h-ratio both over simplicial basins (red) and over lattices (blue), they are visually more evident in red distributions. By Hermite normal form reduction (see Remark 3.3.2), we shrink down the sizes of the bins in the graphs. Therefore, observing smaller spikes in the blue graphs is expected. We observe from the lower right graph in Figure 3.10, that the shrinking of the spikes is not uniform. This is an expected observation since we already know that Hermite normal form reduction changes the expected h-ratios of the distributions. In addition to this, we see that shrinking of the spikes induced by the Hermite normal form reduction smoothens the distribution of the h-ratio. Therefore, considering the h-ratio distribution over lattices is not only plausible mathematically but also statistically. We have already mentioned for both sampled and fully computed cases, that the mean of the h-ratio distribution over the lattices is less than the mean over the simplicial basins. This observation suggests that Hermite normal form reduction has an nontrivial effect on h-ratio distribution. Furthermore, interpreting this observation in terms of circuit polynomials, we conclude that circuit polynomials are less likely to be SOS with respect to the more valid statistics where we consider lattices instead of simplicial basins.

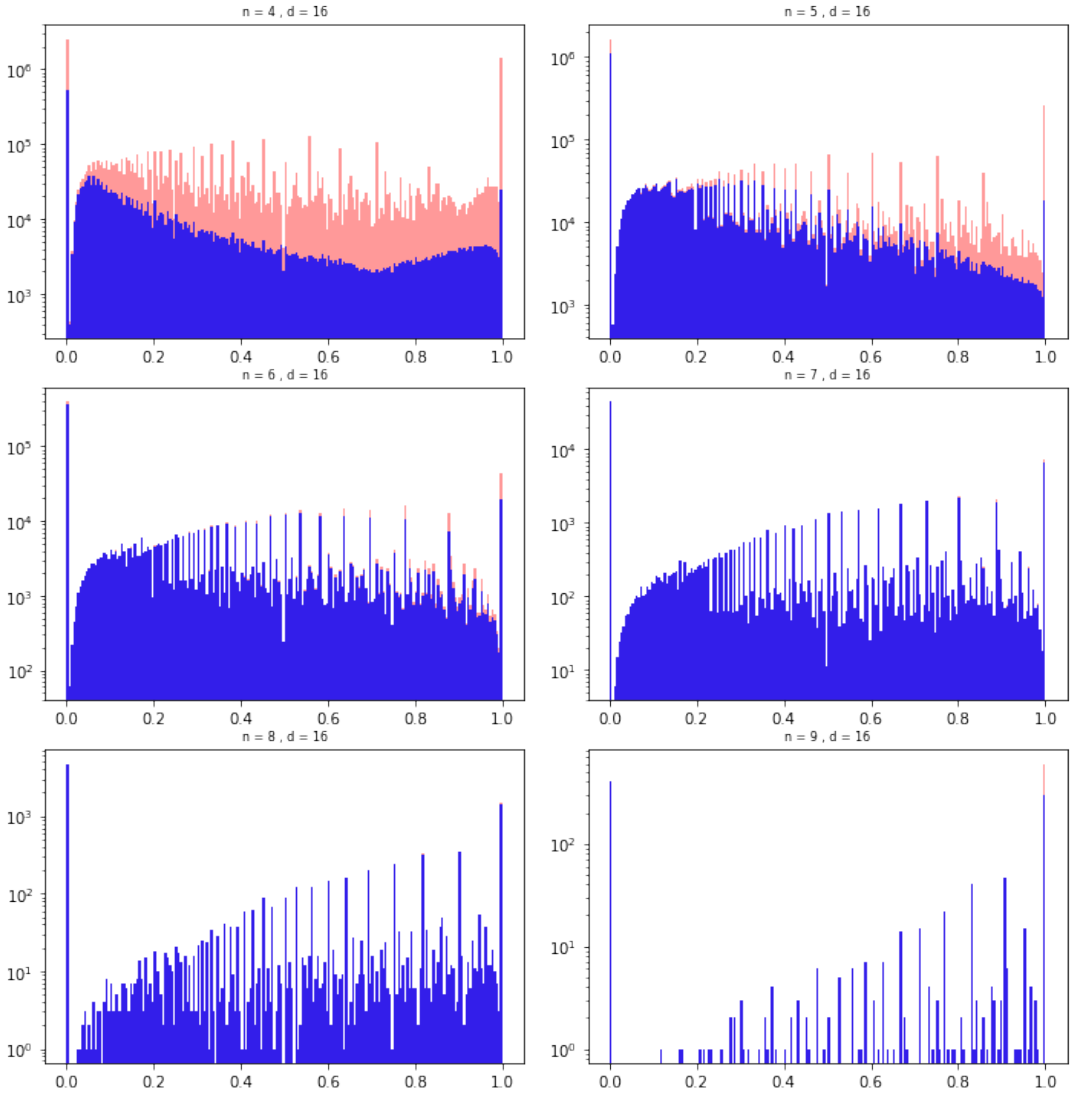

Figure 3.9: The sampled distributions of h-ratio over the n-simplicial basins (red), and over lattices (blue) for the given n and $2d$ in Table 3.2.

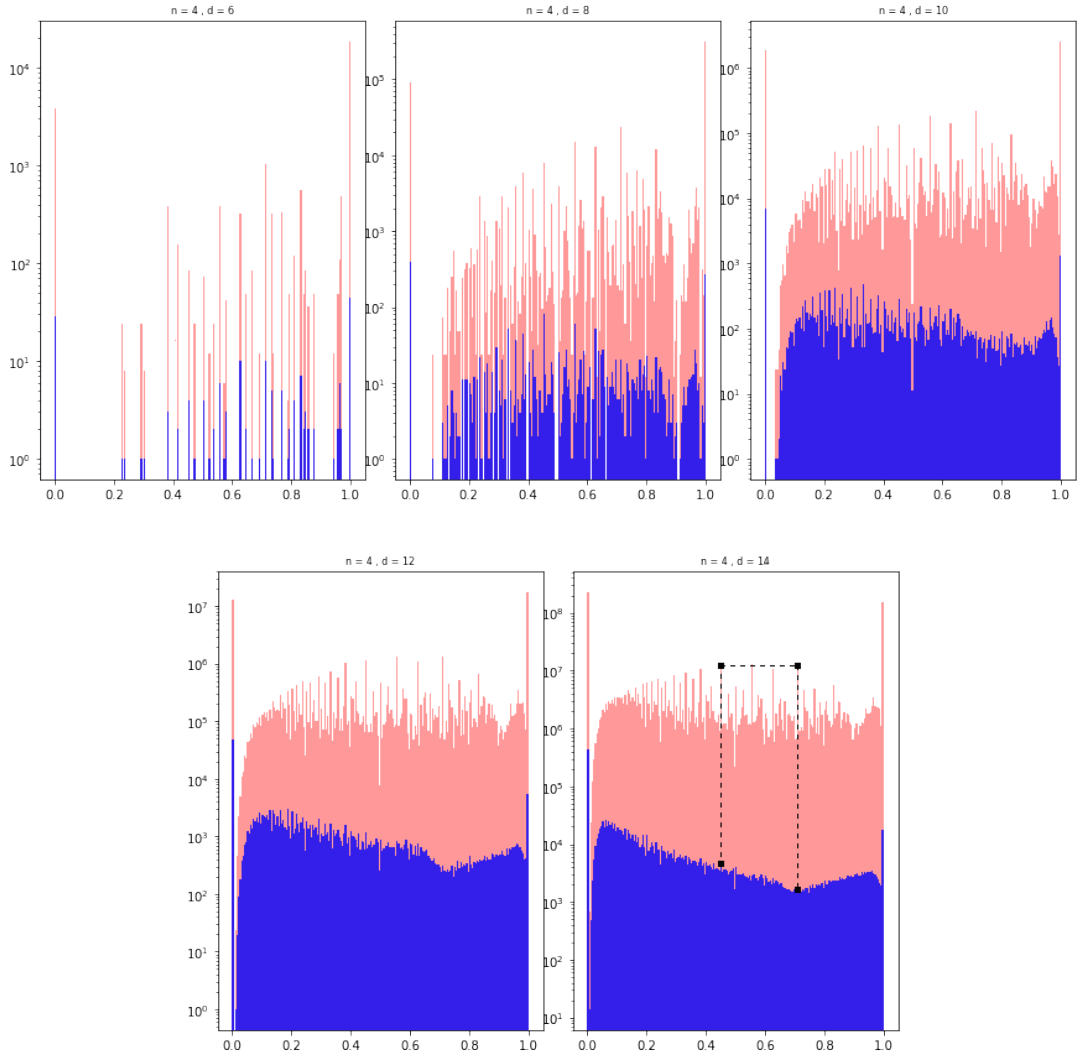

Figure 3.10: The distributions of h-ratio over the 4-simplicial basins (red), and over lattices (blue) for $2d = 6, 8, 10, 12, 14$. On bottom right, we see that two spikes of same height is effected differently from HNF reduction.

Chapter 4

Chemical Reaction Networks

4.1 A General Introduction to Chemical Reaction Network Theory

Chemical reaction network theory, in short CRNT, studies the systems of nonlinear ODEs which describe the behavior of a reaction network. Mathematical models for chemical and biological systems are often based on systems of nonlinear ODEs, e.g. for CRNT, a routine assumption is that the reactions are governed by mass action kinetics, which causes the arising ODEs to be given by polynomial equations. Computing the fixed points of ODEs that underlie chemical and biological systems is an important task, since the fixed points of these ODEs have significant importance in terms of biological dynamics. However, given the nonlinearity of the differential equations and complexity of the investigated systems, the question of whether a relevant system of ODEs has any fixed points is in general hard to answer. Yet, ODE systems that show up in CRNT possess more structure than an arbitrary ODE system due to the stoichiometric restrictions of the reaction network. This is one of the reasons why it was possible to prove various strong results in CRNT. In this section, we give a basic introduction to the CRNT by providing the central definitions and results. As the CRNT is a vast area of study, in Section 4.1 we cover only those parts required to present the results in [FKdWY20]. For a more elaborate discussion of CRNT, we recommend [Gun03] and [Fei19].

4.1.1 Chemical Reaction Networks, Stoichiometry and Mass-Action Kinetics

A chemical reaction network is a finite set of reactions among a finite set of chemical species. In [Fei19, Part 1, Chapter 3], an introduction to the chemical reaction networks

is given in its full formality. As we do not require the full details of the theory, we make a refined introduction focusing on the aspects that are required in Chapter 4.

Definition 4.1.1. A *chemical reaction network* $\mathscr{N} = (\chi, \mathscr{R})$ consists of a set of *species* $\chi := \{X_1, \dots, X_n\}$, and a set of *reactions* $\mathscr{R} := \{R_1, \dots, R_l\}$ of the following form:

$$R_j : \sum_{i=1}^{n} a_{ij} X_i \xrightarrow{\kappa_j} \sum_{i=1}^{n} b_{ij} X_i \tag{4.1.1}$$

where a_{ij}, b_{ij} are nonnegative integers and $\kappa_j \in \mathbb{R}_{\geq 0}$ for all $j = 1, \dots, l$. The coefficients a_{ij}, b_{ij} of the each species any the reaction R_j are called *stoichiometric coefficients*, and they indicate how many units of the species X_i are depleted or produced as the reaction R_j occurs. The rate of the reaction R_j in the network is regulated by the parameter $\kappa_j \in \mathbb{R}_{\geq 0}$ which is called *the reaction rate constant* of R_j. ⎔

Each side of the reaction is called a *complex*, and in particular we call the left hand side the *reactant complex* and the right hand side the *product complex* of the reaction in (4.1.1). If $\mathscr{N} = (\chi, \mathscr{R})$ is a chemical reaction network given as in Definition 4.1.1, then for fixed i and j, the net production of species X_i in reaction R_j is given as $N_{ij} := b_{ij} - a_{ij}$. Subsequently, the *stoichiometric matrix* of this network is defined as $N := (N_{ij}) \in \mathbb{R}^{n \times l}$. The *rank of the network* is the rank of the matrix N, and the *corank of the network* is $n - \operatorname{rank}(N)$.

Example 4.1.2. Consider the chemical reaction network consisting of 2 species X_1, X_2 and the following reactions:

$$\begin{aligned} R_1 &: X_1 \xrightarrow{\kappa_1} 2X_2 \\ R_2 &: 4X_2 \xrightarrow{\kappa_2} 2X_1 \\ R_3 &: X_1 + 2X_2 \xrightarrow{\kappa_3} 4X_2 \end{aligned}$$

Then, the stoichiometric matrix of this chemical reaction network is $N = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \end{bmatrix}$. Rank and corank of the network are 1. ⎔

We denote the concentration of the species X_i with the lower case character x_i, and denote the vector of all concentrations in a reaction network with species $\chi = \{X_1, \dots, X_n\}$ as $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}_{\geq 0}$. We model how $\boldsymbol{x}$ varies as the reactions $R_1, \dots, R_l$ progress simultaneously by constructing a system of ODEs. There are two important ingredients that we use in this construction: the first one is the stoichiometric matrix, and the second one is the choice of kinetics underlying the chemical reaction network. *A reaction*

rate function of R_j is a continuously differentiable function $v_j : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ that models the speed of the reaction with respect to $\boldsymbol{x} = (x_1, \ldots, x_n)$. The set of all reaction rate functions of a reaction network is called as *kinetics* of the reaction network. Of course, not every reaction rate function makes sense in real life circumstances. For example, we assume that a reaction rate function v_j of R_j always satisfies

$$v(\boldsymbol{x}) = 0 \Leftrightarrow x_i = 0 \text{ for some } i \text{ such that } a_{ij} > 0, \tag{4.1.2}$$

because in reality the chemical reaction R_j can occur if and only if all of the species in the reactant complex of R_j exists. In this thesis, we always work with *mass action kinetics*, in which each reaction rate function of R_j for any $j = 1, \ldots, l$ is given as

$$v_j(\boldsymbol{x}) = \kappa_j x_1^{a_{1j}} \ldots x_n^{a_{nj}}. \tag{4.1.3}$$

A chemical reaction system that is regulated by mass action kinetics is called a *mass action system*. The assumption of mass action kinetics states that the reaction rate functions are proportional to the product of the concentrations of the species in the reactant complex.

Let $\mathscr{N} = (\chi, \mathscr{R})$ be a chemical reaction network, and let $\boldsymbol{v(x)} := (v_1(\boldsymbol{x}), \ldots, v_l(\boldsymbol{x}))$ denote the vector of reaction rate functions given by mass action kinetics as in equation (4.1.3). The concentration vector of species, $\boldsymbol{x} = (x_1, \ldots, x_n)$, is time dependent, and we sometimes write $\boldsymbol{x(t)}$ with $t \geq 0$ to stress this (time) dependence. The trajectory of $\boldsymbol{x(t)}$ in $\mathbb{R}_{>0}$ for $t > 0$ is called as the *semi flow of the network*. Furthermore, we let $\dot{\boldsymbol{x}}$ denote the first time derivative of the vector $\boldsymbol{x} = (x_1, \ldots, x_n)$. Next, we consider the following ODE system modeling the vector of concentrations $\boldsymbol{x} = (x_1, \ldots, x_n)$ over time with the initial condition $\boldsymbol{x(0)} \in \mathbb{R}^n_{\geq 0}$:

$$\dot{\boldsymbol{x}} = \boldsymbol{f(x)}, \text{ where } \boldsymbol{f(x)} = N \cdot \boldsymbol{v(x)}. \tag{4.1.4}$$

The function $\boldsymbol{f(x)}$ is sometimes referred to as the *species formation rate function*. Due to the assumption of mass action kinetics, the species formation rate function is a polynomial function from $\mathbb{R}^n_{\geq 0}$ to $\mathbb{R}^n_{\geq 0}$, and the ODE in (4.1.4) is a system consisting n many n-variate polynomial equations. A natural question to ask is, which polynomial systems can arise from a mass action system as described in (4.1.4). This was investigated by Hárs and Tóth under the name of the inverse problem of reaction kinetics in [HT81]. Their result states that a system of polynomial equations

$$\dot{x_i} = f_i(x_1, \ldots, x_n), \text{ where } f_i \in \mathbb{R}[x_1, \ldots, x_n] \text{ for all } i = 1, \ldots, n \tag{4.1.5}$$

arises from a mass action system if and only if each negative monomial of the polynomial f_i contains a nonzero power of x_i for all $i = 1, \ldots, n$.

Example 4.1.3. For the example network given in Example 4.1.2, the vector of reaction rate functions is $\boldsymbol{v}(\boldsymbol{x}) = (\kappa_1 x_1, \kappa_2 x_2^4, k_3 x_1 x_2^2)$. Recall that its stoichiometric matrix is $N = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \end{bmatrix}$, and hence, the ODE associated to this network is given as:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) = N\boldsymbol{v}(\boldsymbol{x}) = \begin{bmatrix} -\kappa_1 x_1 + 2\kappa_2 x_2^4 - \kappa_3 x_1 x_2^2 \\ 2\kappa_1 x_1 - 4\kappa_2 x_2^4 + 2\kappa_3 x_1 x_2^2 \end{bmatrix}.$$

⬡

Let $\mathscr{N} = (\chi, \mathscr{R})$ be a chemical reaction network with the species formation rate function $\boldsymbol{f}(\boldsymbol{x})$ and N be its stoichiometric matrix. Then, the *stoichiometric subspace* S of the reaction network is defined as the linear subspace generated by the columns of N. If the initial species concentration vector in an homogeneous reactor is $\boldsymbol{x}(\boldsymbol{0}) \in \mathbb{R}^n_{\geq 0}$, then the concentration vector at time $t \in \mathbb{R}_{\geq 0}$ is given by $\boldsymbol{x}(\boldsymbol{t})$, and the i-th entry of the vector $\boldsymbol{f}(\boldsymbol{x}(\boldsymbol{t})) \in \mathbb{R}^n$ yields the instantaneous generation rate of the species X_i at time $t \in \mathbb{R}_{\geq 0}$. The stoichiometric structure of the reaction network imposes certain restrictions on how does the solution $\boldsymbol{x}(\boldsymbol{t})$ of the ODE given in equation (4.1.4) evolve over time. In the next lemma, we cite a refined version of a central result on how trajectory of solutions can evolve.

Lemma 4.1.4 (Lemma 3.4.5, [Fei19])**.** Let $\mathscr{N} = (\chi, \mathscr{R})$ be a chemical reaction network with χ and $\mathscr{R}$ given as in (4.1.1), and let $I \subset \mathbb{R}_{\geq 0}$ be an interval such that, for all $t \in I$, $x(t)$ is a solution of (4.1.4). Then for each $t_1, t_2 \in I$ with $t_2 > t_1$, there exists $\boldsymbol{d}(\boldsymbol{x}) \in \mathbb{R}^l_{>0}$ such that

$$\boldsymbol{x}(\boldsymbol{t_2}) - \boldsymbol{x}(\boldsymbol{t_1}) = N\boldsymbol{d}(\boldsymbol{x}),$$

where N is stoichiometric matrix of $\mathscr{N}$. Furthermore, for $j \in [l]$, the j-th component of $\boldsymbol{d}(\boldsymbol{x})$ is given by the integral $\int_{t_1}^{t_2} v_j(\boldsymbol{x}(\boldsymbol{t}))\, dt$, where $v_j(x)$ is the reaction rate function of R_j.

Lemma 4.1.4 implies that, a composition vector $\boldsymbol{x'} = \boldsymbol{x}(\boldsymbol{t'})$ can follow the composition vector $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{t})$ along the solution of (4.1.4) only if $\boldsymbol{x}(\boldsymbol{t'}) - \boldsymbol{x}(\boldsymbol{t})$ lies in the stoichiometric subspace S. We say that $\boldsymbol{x'}$ are $\boldsymbol{x}$ *stoichiometrically compatible* if $\boldsymbol{x'} - \boldsymbol{x} \in S$. The stoichiometric compatibility is an equivalence relation in $\mathbb{R}^n_{\geq 0}$ and $\mathbb{R}^n_{>0}$, and we call the induced equivalence classes as the *stoichiometric compatibility classes* and *positive stoichiometric compatibility classes*, respectively. Regardless of the kinetics, a solution trajectory for (4.1.4) is always confined to an unique stoichiometric compatibility class. In particular, the stoichiometric compatibility class that contains $\boldsymbol{x}(\boldsymbol{0})$ is the set

$(\boldsymbol{x}(\mathbf{0}) + S) \cap \mathbb{R}^n_{\geq 0}$, and the positive stoichiometric compatibility class that contains $\boldsymbol{x}(\mathbf{0})$ is the set $(\boldsymbol{x}(\mathbf{0}) + S) \cap \mathbb{R}^n_{>0}$.

Let $\boldsymbol{w_1} \in \mathbb{R}^n$ be a vector from the orthogonal complement of S, i.e., $\langle \boldsymbol{w_1}, \boldsymbol{x} \rangle = 0$ for any $\boldsymbol{x} \in S$, and let $c_1 := \langle \boldsymbol{w_1}, \boldsymbol{x}(\mathbf{0}) \rangle \in \mathbb{R}$. Note that, for any $\boldsymbol{x}$ in the stoichiometric compatibility class of $\boldsymbol{x}(\mathbf{0})$, i.e., for any $\boldsymbol{x} \in (\boldsymbol{x}(\mathbf{0}) + S) \cap \mathbb{R}^n_{\geq 0}$, the following relation holds:

$$\langle \boldsymbol{w_1}, \boldsymbol{x} \rangle = c_1, \tag{4.1.6}$$

The relation given in (4.1.6) is called an *conservation relation* of the system $\mathscr{N}$. To consider all conservation relation at once, we consider a full rank $(n - r)$ by n matrix W such that $WN = 0$, and let $\boldsymbol{c} := W \cdot \boldsymbol{x}(\mathbf{0}) \in \mathbb{R}^{n-r}$. Similarly, for any $\boldsymbol{x} \in (\boldsymbol{x}(\mathbf{0}) + S) \cap \mathbb{R}^n_{\geq 0}$, it holds that $W \cdot \boldsymbol{x} = \boldsymbol{c}$, and we call

$$\mathscr{P}_{\boldsymbol{c}} := \left\{ \boldsymbol{x} \in \mathbb{R}^n_{\geq 0} \mid W\boldsymbol{x} = \boldsymbol{c} \right\} \tag{4.1.7}$$

as the *stoichiometric compatibility class associated to* $\boldsymbol{c}$. In a like manner, the *positive stoichiometric compatibility class associated to* $\boldsymbol{c}$ is defined as $\mathscr{P}^+_{\boldsymbol{c}} := \mathscr{P}_{\boldsymbol{c}} \cap \mathbb{R}^n_{>0}$. Note that the condition $Wx = c$ forms a system of $n - r$ conversation relations, and we call W *a conversation matrix* of $\mathscr{N}$.

Example 4.1.5. Consider again the example network we introduced in Example 4.1.2, whose stoichiometric matrix is $N = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \end{bmatrix}$. The corank of the system is 1, and $W = \begin{bmatrix} 2 & 1 \end{bmatrix}$ is a 1-by-2 conservation matrix. ⬡

4.1.2 Equilibrium Points and Multistationarity

In this subsection, we first give the definition of equilibrium points for the ODEs induced a chemical reaction network. Then, we introduce the notion of multistationarity, that is the existence of multiple steady states in a system, and discuss a particular approach from [CFMW17] to detect the existence or preclusion of multistationarity.

Given a chemical reaction network $\mathscr{N} = (\chi, \mathscr{R})$ with the species formation rate function $\boldsymbol{f}(\boldsymbol{x})$, we say that a vector of concentrations $\boldsymbol{x}^* \in \mathbb{R}^n_{\geq 0}$ is *an equilibrium point* or *a steady state* of the ODE given by $\mathscr{N}$ if $\boldsymbol{f}(\boldsymbol{x}^*) = \mathbf{0}$. In an equilibrium state, the net production of all species is equal to zero. We are interested in the set of all nonnegative equilibrium points, i.e.,

$$V_f := \left\{ \boldsymbol{x} \in \mathbb{R}^n_{\geq 0} \mid \boldsymbol{f}(\boldsymbol{x}) = \mathbf{0} \right\},$$

because the equilibrium points of a chemical reaction may have biochemical implications. A steady state $\boldsymbol{x}^*$ is called a *boundary steady state* if $\boldsymbol{x}^*$ is in the boundary of the non-negative orthant, i.e. if $\boldsymbol{x}^*$ contains a zero entry. Under mass action kinetics, each entry of $\boldsymbol{f}(\boldsymbol{x}^*)$ is a monomial. In this thesis, the rections that we consider will only yield non constant monomials, and therefore $\mathbf{0}$ is always a boundary steady state. Yet this may not hold for more general reaction networks. Furthermore, an equilibrium point $\boldsymbol{x}^*$ is called a *positive equilbrium point* or *a positive steady state*, if it is not in the boundary. Recall from Section 4.1.1, for the reaction network $\mathscr{N}$, the stoichiometric compatibility divides $\mathbb{R}^n_{\geq 0}$ into equivalence classes, and every stoichiometric compatibility class is given by a vector in $\mathbb{R}^{n-r}$, where r is the rank of $\mathscr{N}$. We are particularly interested in the set of equilibrium points in each positive stoichiometric compatibility class $\mathscr{P}^+_{\boldsymbol{c}}$ for some $\boldsymbol{c} \in \mathbb{R}^{n-r}$, i.e. $V_f \cap \mathscr{P}^+_{\boldsymbol{c}}$. This amounts to finding positive solutions of the polynomial system given by

$$f(\boldsymbol{x}) = 0, \text{ such that } W\boldsymbol{x} = \boldsymbol{c}. \tag{4.1.8}$$

If there exists a $\boldsymbol{c} \in \mathbb{R}^{n-r}$ such that $V_f \cap \mathscr{P}^+_{\boldsymbol{c}}$ contains at least two points, i.e., the system in (4.1.8) has at least two solutions, then we say that the system *enables multistationarity*. besides, if there is only one steady state in each positive stoichiometric compatibility class, then the system is called *monostationary*.

Note that in (4.1.8), $V_f \cap \mathscr{P}^+_{\boldsymbol{c}}$ is described by n equations that define the system $f(\boldsymbol{x}) = 0$ and $n-r$ conservation relations given by the equation $W\boldsymbol{x} = \boldsymbol{c}$. It is redundant to use all of these $2n-r$ relations to describe $V_f \cap \mathscr{P}^+_{\boldsymbol{c}}$, and we follow the systematic way that was described in [CFMW17] to eliminate these redundant equations. Let $W \in \mathbb{R}^{(n-r)\times n}$ be a row reduced conversation matrix associated to the system, let $i_1, \ldots, i_{n-r}$ be the indices of the first nonzero coordinate in each row of W. Then we can express $f_{i_j}(\boldsymbol{x})$,the i_j-th entry of $f(\boldsymbol{x})$, as linear combination of $f(\boldsymbol{x})$'s entries with indices different from $i_1, \ldots, i_{n-r}$, due to the relation arising from the inner product between j-th row W and $f(\boldsymbol{x})$. Consequently, we can disregard the equations $f_{i_1}(\boldsymbol{x}), \ldots, f_{i_{n-r}}(\boldsymbol{x}) = 0$. In order to do so, for $\boldsymbol{c} \in \mathbb{R}^{(n-r)}$, we define the continuously differentiable function $\varphi_c(\boldsymbol{x}) : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n$ as follows:

$$\varphi_c(\boldsymbol{x})_i := \begin{cases} f_i(\boldsymbol{x}) & i \notin \{i_1, \ldots, i_{n-r}\} \\ (W\boldsymbol{x} - \boldsymbol{c})_i & i \in \{i_1, \ldots, i_{n-r}\} \end{cases} \tag{4.1.9}$$

where $f_i(\boldsymbol{x})$ and $(W\boldsymbol{x} - \boldsymbol{c})_i$ denote the i-th entry of the vectors. Since we replaced the redundant equations in $f(\boldsymbol{x})$ with the equations that define $\mathscr{P}_{\boldsymbol{c}}$, the steady states in the

stoichiometric compatibility class $\mathscr{P}_{\boldsymbol{c}}$ is given as:

$$V \cap \mathscr{P}_{\boldsymbol{c}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n_{\geq 0} \mid \varphi_c(\boldsymbol{x}) = 0 \right\}.$$

Therefore, the network that gives rise to $f(\boldsymbol{x})$ enables multistationarity if $\varphi_c(\boldsymbol{x}) = 0$ has at least two solutions for some $\boldsymbol{c} \in \mathbb{R}^{n-r}$. We denote the Jacobian matrix of $\varphi_c(\boldsymbol{x})$ with $M(\boldsymbol{x})$, i.e. the (i,j)-th entry of $M(\boldsymbol{x})$ corresponds to the partial derivative of $\varphi_c(\boldsymbol{x})_i$ with respect to x_j.

Remark 4.1.6. Note that $M(\boldsymbol{x})$ does not depend on the choice of $\boldsymbol{c}$, and the determinant of $M(\boldsymbol{x})$ is a polynomial in $\mathbb{R}[\boldsymbol{x}]$. ⬡

Definition 4.1.7. A steady state $\boldsymbol{x} \in V \cap \mathscr{P}_{\boldsymbol{c}}$ is called *nondegenerate* if $\det(M(\boldsymbol{x})) \neq 0$. ⬡

Example 4.1.8. Consider the chemical reaction network given in Example 4.1.2 that gives rise to the following system of ODEs

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) = N \cdot v(\boldsymbol{x}) = \begin{bmatrix} -\kappa_1 x_1 + 2\kappa_2 x_2^4 - \kappa_3 x_1 x_2^2 \\ 2\kappa_1 x_1 - 4\kappa_2 x_2^4 + 2\kappa_3 x_1 x_2^2 \end{bmatrix}$$

We first recall that rank of N is 1, and the stoichiometric subspace S is generated by the vector $\begin{bmatrix} -1 & 2 \end{bmatrix}$. We compute a conservation matrix W, i.e., a full rank 1 by 2 matrix W such that $WN = 0$ as $W = \begin{bmatrix} 2 & 1 \end{bmatrix}$.

Therefore, each stoichiometry class is defined by the linear equation $2x_1 + x_2 = c$ for some $c \in \mathbb{R}$. W is clearly row reduced, and the first nonzero element in the row is 1, i.e., $i_1 = 1$. Then, we calculate the map $\varphi_c(\boldsymbol{x})$ given in (4.1.9) as follows:

$$\varphi_c(\boldsymbol{x}) = \begin{bmatrix} 2x_1 + x_2 - c \\ 2\kappa_1 x_1 - 4\kappa_2 x_2^4 + 2\kappa_3 x_1 x_2^2 \end{bmatrix}$$

The equation $\varphi_c(\boldsymbol{x})_1$ describes a line for each $c \in \mathbb{R}$, and if we fix the reaction rate constants $\kappa_1, \kappa_2, \kappa_3$, the equation $\varphi_c(\boldsymbol{x})_2 = 0$ describes a degree 4 curve in $\mathbb{R}^2$. We depict the set $V \cap \mathscr{P}_{\boldsymbol{c}}$ for fixed reaction rate constants $\kappa_1 = 3, \kappa_2 = 1, \kappa_3 = 2$, and for $c = 1, 2, 3$ in Figure 4.1. ⬡

In [CFMW17], the authors point out that, for a particular class of networks, the existence of multistationarity can be decided by studying the sign of a relevant polynomial, which we describe in Theorem 4.1.10. Before stating Theorem 4.1.10, we define a technical property for reaction networks, which is going to be a necessary condition for Theorem 4.1.10.

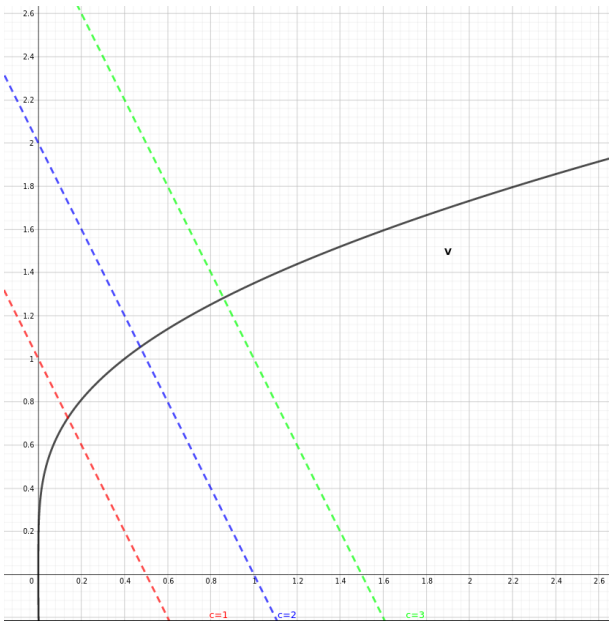

Figure 4.1: The black solid curve is the set of steady states V for the running example we introduced in Example 4.1.2. The red, blue, and green dashed lines show the stoichiometric compatibility classes for $c = 1$, $c = 2$ and $c = 3$ respectively.

Definition 4.1.9. A chemical reaction network $\mathscr{N} = (\chi, \mathscr{R})$ with the stoichiometric subspace $S \subset \mathbb{R}^n$, is called *conservative* if $S^\perp \cap \mathbb{R}^n_{>0} \neq 0$, where $^\perp$ denotes the orthogonal complement of S in $\mathbb{R}^n$. ⎔

We note that throughout this thesis we only work with conservative networks.

Theorem 4.1.10 (Theorem 1, [CFMW17])**.** Let $\mathscr{N} = (\chi, \mathscr{R})$ be a conservative chemical reaction network with stoichiometric matrix $N \in \mathbb{R}^{n\times l}$ of rank r and the reaction rate function $v(\boldsymbol{x})$ that satisfies (4.1.2). Furthermore, let $\mathscr{P}_{\boldsymbol{c}}$ be a nonempty stoichiometric compatibility class without any boundary equilibrium point where $\boldsymbol{c} \in \mathbb{R}^{n-r}$ and let $M(\boldsymbol{x})$ denote the Jacobian matrix of $\varphi_c(\boldsymbol{x})$. Then, the following statements hold:

(a) If $\operatorname{sign}(\det(M(\boldsymbol{x}))) = (-1)^r$ for all $\boldsymbol{x} \in V \cap \mathscr{P}^+_{\boldsymbol{c}}$, then there is exactly one positive equilibrium in $\mathscr{P}_{\boldsymbol{c}}$ and it is nondegenerate.

(b) If $\operatorname{sign}(\det(M(\boldsymbol{x}))) = (-1)^{r+1}$ for some $\boldsymbol{x} \in V \cap \mathscr{P}^+_{\boldsymbol{c}}$, then there are at least two positive equilibria in $\mathscr{P}_{\boldsymbol{c}}$, and at least one of them is nondegenerate. If all positive equilibria in $\mathscr{P}_{\boldsymbol{c}}$ are nondegenerate, then the number of positive equilibria is $2k + 1$ for some positive integer k.

Remark 4.1.11. In [CFMW17, Theorem 1], the authors prove Theorem 4.1.10 for dissipative networks, which we do not define here. However, we note that dissipativity is

a weaker assumption than conservativity, since all conservative networks are dissipative, see [CFMW17, Supplementary Information, Section 3.2]. ⬡

Theorem 4.1.10 itself is not directly very useful for detecting multistationarity because it requires us to compute $\boldsymbol{x} \in V \cap \mathscr{P}_{\boldsymbol{c}}^{+}$. In order to utilize this theorem without computing the set of steady states, the authors point out a corollary of this theorem in [CFMW17, Corollary 2] that utilizes parameterization of the positive equilibrium points. A positive parameterization of the set of positive equilibria is a surjective function

$$\begin{aligned} \Phi : \mathbb{R}^{m}_{>0} &\to V \cap \mathbb{R}^{n}_{>0} \\ \widetilde{\boldsymbol{x}} = (\widetilde{x_1}, \ldots, \widetilde{x_m}) &\mapsto (\Phi_1(\widetilde{\boldsymbol{x}}), \ldots, \Phi_n(\widetilde{\boldsymbol{x}}))\,, \end{aligned} \tag{4.1.10}$$

where $\widetilde{\boldsymbol{x}}$ is the vector of free variables, whose entries form a cardinality m subset of the variables $x_1, \ldots, x_n$ for some $m < n$. We note that $x_1, \ldots, x_n$ are positive provided that $\widetilde{\boldsymbol{x}}$ is positive. Given a chemical reaction network $\mathscr{N} = (\chi, \mathscr{R})$ as in (4.1.1), under the assumption of mass action kinetics, the equation $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$ results in $\operatorname{rank} \mathscr{N} < n$ polynomial equations in n unknowns, which generically yields an $n - r$ dimensional parameterization. If such a positive parameterization exists for $m = n - r$, then $x_1, \ldots, x_n$ can be written as functions of $\widetilde{\boldsymbol{x}}$ at equilibrium points as $x_i = \Phi_i(\widetilde{\boldsymbol{x}})$ for all $i = 1, \ldots, n$. As we would like to use same parameterization in all compatibility classes, we avoid using the conversation relations to come up with these positive parameterizations. With this in mind, we note that a positive equilibrium given by $\Phi(\widetilde{\boldsymbol{x}})$ for some $\widetilde{\boldsymbol{x}}$ belongs to the stoichiometric compatibility class $\mathscr{P}_{\boldsymbol{c}}$ such that $c = W\Phi(\widetilde{\boldsymbol{x}})$. Furthermore, the positive solutions of (4.1.8) can be restated as

$$V \cap \mathscr{P}_{\boldsymbol{c}}^{+} = \{\Phi(\widetilde{\boldsymbol{x}}) \mid \widetilde{\boldsymbol{x}} \in \mathbb{R}^{m}_{>0} \text{ and } W\Phi(\widetilde{\boldsymbol{x}}) = \boldsymbol{c}\}\,. \tag{4.1.11}$$

In order to express the next theorem more clearly, let $p(\widetilde{\boldsymbol{x}})$ denote the evaluation of the determinant of $M(\boldsymbol{x})$ at $\Phi(\widetilde{\boldsymbol{x}})$:

$$p(\widetilde{\boldsymbol{x}}) := \det(M(\Phi(\widetilde{\boldsymbol{x}})), \tag{4.1.12}$$

for $\widetilde{\boldsymbol{x}} \in \mathbb{R}^{m}_{>0}$. We note that $p(\widetilde{\boldsymbol{x}})$ does not depend on the choice of $\boldsymbol{c}$ due to Remark 4.1.6.

Theorem 4.1.12 (Corollary 2, [CFMW17])**.** Let $\mathscr{N} = (\chi, \mathscr{R})$ be a chemical reaction network that satisfies the assumptions of Theorem 4.1.10. Furthermore, assume that there exists a positive parameterization $p(\widetilde{\boldsymbol{x}})$ of the set of positive equilibria. Then, the following statements hold:

(a) If $\operatorname{sign}(p(\widetilde{\boldsymbol{x}})) = (-1)^r$ for all $\widetilde{\boldsymbol{x}} \in \mathbb{R}^{m}_{>0}$, then there is exactly one positive equilibrium in each $\mathscr{P}_{\boldsymbol{c}}$ with $\mathscr{P}_{\boldsymbol{c}}^{+}$, and this equilibrium point is nondegenerate.

(b) If $\text{sign}(p(\widetilde{\boldsymbol{x}})) = (-1)^{r+1}$ for some $\widetilde{\boldsymbol{x}} \in \mathbb{R}^m_{>0}$, then there are at least two positive equilibria in $\mathscr{P}_{\boldsymbol{c}}$ for $c = W\Phi(\widetilde{\boldsymbol{x}})$. Furthermore, at least one of the equilibrium points is nondegenerate. If all positive equilibria in $\mathscr{P}_{\boldsymbol{c}}$ are nondegenerate, then the number of positive equilibria is $2k+1$ for some positive integer k.

The Theorem 4.1.12 will be the main machinery as we study the multistationarity and monostationarity of the chemical reaction networks.

Example 4.1.13. Consider again the running example that we introduced in Example 4.1.2. In Example 4.1.8 we computed map $\varphi_c(\boldsymbol{x})$ associated to this example. The Jacobian matrix associated to the map $\varphi_c(\boldsymbol{x})$ is given as:

$$M(\boldsymbol{x}) = \begin{bmatrix} 2 & 1 \\ 2\kappa_1 + 2\kappa_3 x_2^2 & -16\kappa_2 x_2^3 + 4\kappa_3 x_1 x_2 \end{bmatrix}$$

The determinant of $M(\boldsymbol{x})$ is $-32\kappa_2 x_2^3 + 8\kappa_3 x_1 x_2 - 2\kappa_1 - 2\kappa_3 x_2^2$. The equation $\varphi_c(\boldsymbol{x})_2 = 0$ yields that $x_1 = \frac{2\kappa_2 x_2^4}{\kappa_1 + \kappa_3 x_2^2}$, and we can parameterize the positive steady states with $\widetilde{\boldsymbol{x}} = x_2$ as follows:

$$\Phi(\widetilde{\boldsymbol{x}}) = \left(\frac{2\kappa_2 x_2^4}{\kappa_1 + \kappa_3 x_2^2}, x_2 \right)$$

Note that since the corank of the system is 1, we only require one parameter for this positive parameterization. By substituting the x_1 and x_2 in the determinant with the parameterization given by $\Phi(\widetilde{\boldsymbol{x}})$, we compute that

$$\begin{aligned} p(\widetilde{\boldsymbol{x}}) &= -32\kappa_2 x_2^3 + 8\kappa_3 \frac{2\kappa_2 x_2^4}{\kappa_1 + \kappa_3 x_2^2} x_2 - 2\kappa_1 - 2\kappa_3 x_2^2 \\ &= \frac{-32\kappa_1\kappa_2 x_2^3 - 16\kappa_2\kappa_3 x_2^5 - 2\kappa_1^2 - 2\kappa_1\kappa_3 x_2^2 - 2\kappa_1\kappa_3 x_2^2 - 2\kappa_3^3 x^4}{\kappa_1 + \kappa_3 x_2^2} \end{aligned}$$

Note that the sign of $p(\widetilde{\boldsymbol{x}})$ is negative for all $x_2 > 0$, which is equal to $(-1)^r$ since the rank of the system is 1. Therefore, by Theorem 4.1.12 part (a) the system is monostationary for all $\kappa_1, \kappa_2, \kappa_3$. This was also evident in the Figure 4.1: since any stoichiometric compatibility class, which is in fact given by a line parallel to the dashed lines, intersects V in a unique point in $\mathbb{R}^2_{>0}$. ⬡

The running example we followed throughout this section was unrealistically small for real life cases. We conclude this section by giving a more realistic example, and go over our main strategy for the certification of monostationarity and multistationarity.

Example 4.1.14. We consider a chemical reaction network system that arises from *hybrid histidine kinase system*, which is a 2-component regulatory system. Histidine kinase has

two sites of phosphorylation which are ordered, and its regulatory role depends on which site is phosphorylated, see [KFCS15] for more on hybrid histidine kinase system. The chemical reaction network arising form the system is given as follows:

$$\begin{aligned} HK_{00} &\xrightarrow{\kappa_1} HK_{p0} \xrightarrow{\kappa_2} HK_{0p} \xrightarrow{\kappa_3} HK_{pp} \\ HK_{0p} + RR &\xrightarrow{\kappa_4} HK_{00} + RR_p \\ HK_{pp} + RR &\xrightarrow{\kappa_5} HK_{p0} + RR_p \\ RR_p &\xrightarrow{\kappa_6} HK_{pp} \end{aligned}$$

HK_{00} and HK_{pp} denote the histidine kinase with zero and two phosphorylated sites, respectively. Also, HK_{p0} and HK_{0p} denote the histidine kinase with its first and the second site phosphorylated, respectively. RR is the response regulator protein, and it may be phosporylated by histidine kinase and form RR_p. We use the notation $X_1 = HK_{00}, X_2 = HK_{p0}, X_3 = HK_{0p}, X_4 = HK_{pp}, X_5 = RR, X_6 = RR_p$ to denote the species, and lower case x_i to denote the amounts of each species. Then the stoichiometric matrix is given as

$$N = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}.$$

The vector of reaction rate functions is $v(\boldsymbol{x}) = (\kappa_1 x_1, \kappa_2 x_2, \kappa_3 x_3, \kappa_4 x_3 x_5, \kappa_5 x_4 x_5, \kappa_6 x_6)$, and the species formation rate is given as

$$f(\boldsymbol{x}) = N \cdot v(\boldsymbol{x}) = \begin{bmatrix} -\kappa_1 x_1 + \kappa_4 x_3 x_5 \\ \kappa_1 x_1 - \kappa_2 x_2 + \kappa_5 x_4 x_5 \\ \kappa_2 x_2 - \kappa_3 x_3 - \kappa_4 x_3 x_5 \\ -\kappa_4 x_3 x_5 - \kappa_5 x_4 x_5 + \kappa_6 x_6 \\ \kappa_4 x_3 x_5 + \kappa_5 x_4 x_5 - \kappa_6 x_6 \end{bmatrix}.$$

The network is conservative since all ones vector $\mathbf{1} \in \mathbb{R}^6$ is in $\mathrm{im}(N)^\perp$, and therefore it is dissipative. Note that the rank of N is 4, and consider the rank 2 matrix

$$W = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

which satisfies $WN = 0$. Each vector $\boldsymbol{c} = (c_1, c_2) \in \mathbb{R}^2$ defines a stoichiometric compatibility class, and W gives rise to the conversation relations $x_1 + x_2 + x_3 + x_4 = c_1$ and $x_5 + x_6 = c_2$.

In order to compute $V \cap \mathscr{P}_{\boldsymbol{c}}^+$, it is redundant to consider all of the equations defined by $f(\boldsymbol{x}) = \mathbf{0}$ and conversation relations. To avoid redundancy, we define $\varphi_c(\boldsymbol{x})$ as it was described in (4.1.9):

$$\varphi_c(\boldsymbol{x}) = \begin{bmatrix} x_1 + x_2 + x_3 + x_4 - c_1 \\ \kappa_1 x_1 - \kappa_2 x_2 + \kappa_5 x_4 x_5 \\ \kappa_2 x_2 - \kappa_3 x_3 - \kappa_4 x_3 x_5 \\ x_5 + x_6 - c_2 \\ \kappa_4 x_3 x_5 + \kappa_5 x_4 x_5 - \kappa_6 x_6 \end{bmatrix}.$$

The Jacobian of $\varphi_c(\boldsymbol{x})$ is

$$M(\boldsymbol{x}) = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ \kappa_1 & -\kappa_2 & 0 & \kappa_5 x_5 & \kappa_5 x_4 & 0 \\ 0 & \kappa_2 & -\kappa_3 - \kappa_4 x_5 & 0 & -\kappa_4 x_3 & 0 \\ 0 & 0 & \kappa_3 & -\kappa_5 x_5 & -\kappa_5 x_4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & \kappa_4 x_5 & \kappa_5 x_5 & \kappa_4 x_3 + \kappa_5 x_4 & -\kappa_6 \end{bmatrix},$$

and the determinant of $M(\boldsymbol{x})$ is given as

$$\begin{aligned} M(\boldsymbol{x}) =& \kappa_2\kappa_4\kappa_5(\kappa_1 - \kappa_3)x_3x_5 + \kappa_1\kappa_2\kappa_4\kappa_5 x_4 x_5 + \kappa_4\kappa_5\kappa_6(\kappa_1 + \kappa_2)x_5^2 \\ &+ \kappa_1\kappa_2\kappa_3\kappa_4 x_3 + \kappa_1\kappa_2\kappa_3\kappa_5 x_4 + \kappa_1\kappa_5\kappa_6(\kappa_3 + \kappa_2)x_5 + \kappa_1\kappa_2\kappa_3\kappa_6. \end{aligned}$$

It is clear that the sign of the determinant is positive, if $\kappa_1 \geq \kappa_3$. Therefore, Theorem 4.1.10 implies that system has a unique equilibrium point in each stoichiometric compatibility class if $\kappa_1 \geq \kappa_3$. The sign of the determinant is not clear for $\kappa_1 < \kappa_3$, and we would like to study its sign through its Newton polytope which is hard to achieve in dimension 6. So, we employ a positive parameterization $\Phi(x_4, x_5) : \mathbb{R}_{>0}^{(6-4)} \to V \cap \mathbb{R}_{>0}^6$

$$\Phi(x_4, x_5) = \left(\frac{\kappa_4\kappa_5 x_4 x_5^2}{\kappa_1\kappa_3}, \frac{\kappa_5(\kappa_4 x_5 + \kappa_3)x_4 x_5}{\kappa_2\kappa_3}, \frac{\kappa_5 x_4 x_5}{\kappa_3}, x_4, x_5, \frac{\kappa_5(\kappa_4 x_5 + \kappa_3)x_4 x_5}{\kappa_3\kappa_6} \right)$$

by solving the equations $f_1(\boldsymbol{x}) = f_2(\boldsymbol{x}) = f_3(\boldsymbol{x}) = f_1(\boldsymbol{x})$ for x_1, x_2, x_3 and x_4 respectively.

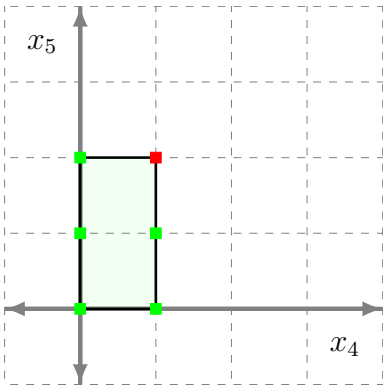

Figure 4.2: Light green rectangle represents is the Newton polytope of $p(x_4, x_5)$ from Example 4.1.14. The lattice point marked with red square corresponds to the term with the negative coefficient, and the ones marked with green correspond to the terms with positive coefficients.

Under the positive parameterization $\Phi(x_4, x_5)$, the determinant of the Jacobian becomes:

$$p(x_4, x_5) = \frac{1}{\kappa_3}\left(\kappa_2\kappa_4\kappa_5^2(\kappa_1 - \kappa_3)x_4x_5^2 + (\kappa_1 + \kappa_2)\kappa_3\kappa_4\kappa_5\kappa_6x_5^2 + 2\kappa_1\kappa_2\kappa_3\kappa_4\kappa_5x_4x_5 + (\kappa_2 + \kappa_3)\kappa_1\kappa_3\kappa_5\kappa_6x_5 + \kappa_1\kappa_2\kappa_3^2\kappa_5x_4 + k_1\kappa_2\kappa_3^2\kappa_6\right).$$

The Newton polytope of $p(x_4, x_5)$ is two dimensional, and it is easy to see from Figure 4.2 that the exponent of the term corresponding to the monomial $x_4x_5^2$ is a vertex. Furthermore, the coefficient of this term is negative if $\kappa_1 < \kappa_3$. This implies that $p(x_4, x_5)$ becomes negative for some choice of $x_4, x_5 \in \mathbb{R}^2_{\geq 0}$, see Proposition 4.2.6 or Proposition 4.2.7. Therefore Theorem 4.1.12 implies that the system enables multistationarity if $\kappa_1 < \kappa_3$. ⬡

4.2 Case Study: Phosphorylation Cycle

In this section, we focus on a specific reaction network that arises from a simple model of phosphorylation and dephosphorylation, which is a significant network for various reasons. First, phosphorylation processes are central in the modulation of cell communication, activities and responses, as, for example, phosphorylation affects about 30% of all proteins in human body [Coh89]. Second, this model is a building block of the MAPK cascade, i.e., mitogen activated protein kinase cascade, which are signaling pathways that regulate a wide variety of stimulated cellular activities [HF96, QNKS07, HR17]. Third, in addition to the biological relevance of this system, this network has become the *model*

model (like the model organisms in biology), where new techniques, strategies, and approaches are tested. The reaction network arising from this system is large enough for hands-on approaches to fail, but small enough to challenge the development of new mathematics. Furthermore, dynamical properties of the ODE system of this network might be lifted to more complex networks related to it.

The phosphorylation process may occur in multiple sites of the protein, so the n-site phosphorylation cycle models the case where protein has n possible sites for phosphorylation and dephosphorylation. The reaction network for the n-site case is given as follows:

$$\begin{aligned} S_0 + E &\rightleftharpoons ES_0 \longrightarrow \ldots \rightleftharpoons ES_{n-1} \longrightarrow S_n + E \\ S_n + F &\rightleftharpoons FS_{n-1} \longrightarrow \ldots \rightleftharpoons FS_1 \longrightarrow S_0 + F \end{aligned} \tag{4.2.1}$$

where S is the substrate with $n > 1$ phosphorylation sites, S_i denotes the phosphoforms of S with k phosphorylated sites, E and F denote the kinase and phosphatase enzymes, see [WS08, TG09b, FHC14]. (4.2.2) is also an example of a post-translational modification network [TG09a, FW13, CS18], a MESSI system [PMD18], and a network with toric steady states [MDSC12]. Currently, it is known that the number of positive steady states within a linear invariant subspace is either one or three, if all positive steady states are nondegenerate [WS08, MHK04]. It has also been shown that there are choices of parameters for which there are two asymptotically stable steady states and one unstable steady state [HR15], see also [TF20]. It is currently unknown whether it admits Hopf bifurcations or periodic solutions [CFM20]. An elaborate discussion of the 2-site phosphorylation cycle has been given recently in [FKdWY20], in which a parameterization of the boundary between monostationarity and multistationarity regions have been provided for the case of 2-site. In Section 4.3, we present the results established in [FKdWY20], and particularly in Section 4.3.2, we present a novel method for checking monostationarity that utilizes circuit polynomials as nonnegativity certificates. We extend this method to 3-site model in Section 4.4, and we expect that this strategy can be extendable not only to the higher cases, but also to similar systems arising in molecular biology.

4.2.1 An Overview of the 2-site Phosphorylation Cycle

Especially in Section 4.3, the reaction network we consider will consist of a substrate S that has two phosphorylation sites. Therefore, we initially introduce the reaction network that arises from 2-site phosphorylation cycle, to keep the introduction simple and more accessible. However, later on in Section 4.4, we will also introduce and work with substrates with more than 2 phosphorylation sites. Phosphorylation occurs distributively in an ordered manner, such that one of the sites is always phosphorylated first. We

denote the three phosphoforms of S with $0, 1, 2$ phosphorylated sites by S_0, S_1, S_2 respectively, and assume that a kinase E and a phosphatase F mediate the phosphorylation and dephosphorylation of S respectively. This gives rise to the following chemical reaction network ([WS08, CM14]):

$$\begin{aligned} S_0 + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} ES_0 \xrightarrow{\kappa_3} S_1 + E \underset{\kappa_8}{\overset{\kappa_7}{\rightleftharpoons}} ES_1 \xrightarrow{\kappa_9} S_2 + E \\ S_2 + F \underset{\kappa_{11}}{\overset{\kappa_{10}}{\rightleftharpoons}} FS_2 \xrightarrow{\kappa_{12}} S_1 + F \underset{\kappa_5}{\overset{\kappa_4}{\rightleftharpoons}} FS_1 \xrightarrow{\kappa_6} S_0 + F. \end{aligned} \tag{4.2.2}$$

As it was discussed in Section 4.1, the assumption of mass action kinetics implies that the evolution of the concentration of the species of the network (4.2.2) over time is modeled by a system of autonomous ODEs in $\mathbb{R}^9_{\geq 0}$. We denote the concentrations of the species by $x_1 = [E], x_2 = [F]$, $x_3 = [S_0]$, $x_4 = [S_1]$, $x_5 = [S_2]$, $x_6 = [ES_0]$, $x_7 = [FS_1]$, $x_8 = [ES_1]$, $x_9 = [FS_2]$. Under mass-action kinetics, the ODE system modeling the concentrations of the nine species in the network (4.2.2) over time t is

$$\begin{aligned} \tfrac{dx_1}{dt} &= -\kappa_1 x_1 x_3 - \kappa_7 x_1 x_4 + \kappa_2 x_6 + \kappa_3 x_6 + \kappa_8 x_8 + \kappa_9 x_8 & \tfrac{dx_6}{dt} &= \kappa_1 x_1 x_3 - \kappa_2 x_6 - \kappa_3 x_6 \\ \tfrac{dx_2}{dt} &= -\kappa_4 x_2 x_4 - \kappa_{10} x_2 x_5 + \kappa_5 x_7 + \kappa_6 x_7 + \kappa_{11} x_9 + \kappa_{12} x_9 & \tfrac{dx_7}{dt} &= \kappa_4 x_2 x_4 - \kappa_5 x_7 - \kappa_6 x_7 \\ \tfrac{dx_3}{dt} &= -\kappa_1 x_1 x_3 + \kappa_2 x_6 + \kappa_6 x_7 & \tfrac{dx_8}{dt} &= \kappa_7 x_1 x_4 - \kappa_8 x_8 - \kappa_9 x_8 \\ \tfrac{dx_4}{dt} &= -\kappa_4 x_2 x_4 - \kappa_7 x_1 x_4 + \kappa_3 x_6 + \kappa_5 x_7 + \kappa_8 x_8 + \kappa_{12} x_9 & \tfrac{dx_9}{dt} &= \kappa_{10} x_2 x_5 - \kappa_{11} x_9 - \kappa_{12} x_9 \\ \tfrac{dx_5}{dt} &= -\kappa_{10} x_2 x_5 + \kappa_9 x_8 + \kappa_{11} x_9, \end{aligned} \tag{4.2.3}$$

where $x_i = x_i(t)$, [CM14]. This is a polynomial ODE system whose coefficients are given by the reaction rate constants $\kappa_1, \ldots, \kappa_{12} > 0$. The positive and nonnegative orthants of $\mathbb{R}^9$ are forward invariant by the trajectories of this system as it is the case for all mass-action systems, see Section 4.1.1.

The stoichiometric matrix of the system is

$$N = \begin{bmatrix} -1 & 1 & 1 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & 0 & -1 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & -0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{bmatrix}. \tag{4.2.4}$$

The vector of reaction rates is given as

$$v(\boldsymbol{x}) = (-\kappa_1 x_1 x_3, \kappa_2 x_6, \kappa_3 x_6, \kappa_4 x_2 x_4, \kappa_5 x_7, \kappa_6 x_7, \kappa_7 x_1 x_4, \kappa_8 x_8, \kappa_9 x_8, \kappa_{10} x_2 x_5, \kappa_{11} x_9, \kappa_{12} x_9),$$

and the ODE in (4.2.3) can be expressed as

$$f(\boldsymbol{x}) = N \cdot v(\boldsymbol{x}) = \begin{bmatrix} -\kappa_1 x_1 x_3 - \kappa_7 x_1 x_4 + \kappa_2 x_6 + \kappa_3 x_6 + \kappa_8 x_8 + \kappa_9 x_8 \\ -\kappa_4 x_2 x_4 - \kappa_{10} x_2 x_5 + \kappa_5 x_7 + \kappa_6 x_7 + \kappa_{11} x_9 + \kappa_{12} x_9 \\ -\kappa_1 x_1 x_3 + \kappa_2 x_6 + \kappa_6 x_7 \\ -\kappa_4 x_2 x_4 - \kappa_7 x_1 x_4 + \kappa_3 x_6 + \kappa_5 x_7 + \kappa_8 x_8 + \kappa_{12} x_9 \\ -\kappa_{10} x_2 x_5 + \kappa_9 x_8 + \kappa_{11} x_9 \\ \kappa_1 x_1 x_3 - \kappa_2 x_6 - \kappa_3 x_6 \\ \kappa_4 x_2 x_4 - \kappa_5 x_7 - \kappa_6 x_7 \\ \kappa_7 x_1 x_4 - \kappa_8 x_8 - \kappa_9 x_8 \\ \kappa_{10} x_2 x_5 - \kappa_{11} x_9 - \kappa_{12} x_9 \end{bmatrix}. \tag{4.2.5}$$

The rank of N is 6, and we write a row reduced matrix W whose rows form a basis of $\mathrm{im}(N)^\perp$

$$W = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{4.2.6}$$

The network is conservative, since

$$\begin{pmatrix} -1 & 1 & 2 \end{pmatrix} \cdot W = \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 1 & 3 & 1 & 3 \end{pmatrix} \in \mathbb{R}^9_{>0}, \tag{4.2.7}$$

and consequently it is dissipative. The matrix W gives rise to three conservation relations, which are independent of κ_i.

It follows that, the stoichiometric compatibility class that corresponds to a stoichiometric compatibility class $\mathscr{P}_{\boldsymbol{c}}$ is six dimensional subspace in $\mathbb{R}^9$, defined by the equations

$$x_1 + x_6 + x_8 = c_1, \quad x_2 + x_7 + x_9 = c_2, \quad x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 = c_3, \tag{4.2.8}$$

subject to $x_i \geq 0$ for $i = 1, \ldots, 9$. Here, c_1, c_2, c_3 stand for the total amounts of kinase E, phosphatase F and substrate S respectively. The steady states of the network are the solutions to the system of polynomial equations given by setting $f(\boldsymbol{x}) = 0$ in (4.2.5). If we want to compute the steady states in a given compatibility class $\mathscr{P}_{\boldsymbol{c}}$, then three of these equations are redundant, for example the ones for f_1, f_2, f_3 can be removed. The remaining six equations together with the equations in (4.2.8) form the *steady state system*, which has variables $x_1, \ldots, x_9$ and parameters $\kappa_1, \ldots, \kappa_{12}$, c_1, c_2, c_3, all of which are assumed to

be positive. We derive the map $\varphi_c(\boldsymbol{x})$ from the entries of $f(\boldsymbol{x})$ and conservation relations, as described in (4.1.9). Note that the first nonzero entries in the rows of W are $i_1 = 1$, $i_2 = 2$ and $i_3 = 3$, so $\varphi_c(\boldsymbol{x})$ is given as

$$\varphi_c(\boldsymbol{x}) = \begin{bmatrix} -x_1 + x_6 + x_8 - c_1 \\ x_2 + x_7 + x_9 - c_2 \\ x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 - c_3 \\ -\kappa_4 x_2 x_4 - \kappa_7 x_1 x_4 + \kappa_3 x_6 + \kappa_5 x_7 + \kappa_8 x_8 + \kappa_{12} x_9 \\ -\kappa_{10} x_2 x_5 + \kappa_9 x_8 + \kappa_{11} x_9 \\ \kappa_1 x_1 x_3 - \kappa_2 x_6 - \kappa_3 x_6 \\ \kappa_4 x_2 x_4 - \kappa_5 x_7 - \kappa_6 x_7 \\ \kappa_7 x_1 x_4 - \kappa_8 x_8 - \kappa_9 x_8 \\ \kappa_{10} x_2 x_5 - \kappa_{11} x_9 - \kappa_{12} x_9 \end{bmatrix}. \tag{4.2.9}$$

The steady states in the stoichiometric compatibility class $\mathscr{P}_{\boldsymbol{c}}$ are then given as

$$V \cap \mathscr{P}_{\boldsymbol{c}} = \left\{ \boldsymbol{x} \in \mathbb{R}^n_{\geq 0} \mid \varphi_c(\boldsymbol{x}) = 0 \right\}.$$

Following the notation in Section 4.1, we denote the Jacobian of the map $\varphi_c(\boldsymbol{x})$ with $M(\boldsymbol{x})$, and with Theorem 4.1.10 in mind, we are interested in the sign of $\det(M(\boldsymbol{x}))$. Note that $\det(M(\boldsymbol{x}))$ is a polynomial in $\mathbb{R}[x_1, \ldots, x_9]$,but the last six indeterminate values have a positive parameterization in terms of the first three. Indeed, by solving solving the equations $f_i = 0$ for $i = 4, \ldots, 9$, we get the positive parameterization $\Phi(x_1, x_2, x_3) : \mathbb{R}^3_{>0} \to V \cap \mathbb{R}^9_{>0}$ given by:

$$\begin{aligned} \Phi(x_1, x_2, x_3) = \Bigg(& x_1, x_2, x_3, \frac{\kappa_1\kappa_3(\kappa_5+\kappa_6)x_1x_3}{(\kappa_2+\kappa_3)\kappa_4\kappa_6x_2}, \frac{\kappa_1\kappa_3(\kappa_5+\kappa_6)\kappa_7\kappa_9(\kappa_{11}+\kappa_{12})x_1^2x_3}{(\kappa_2+\kappa_3)\kappa_4\kappa_6(\kappa_8+\kappa_9)\kappa_{10}\kappa_{12}x_2^2}, \frac{\kappa_1x_1x_3}{\kappa_2+\kappa_3}, \\ & \frac{\kappa_1\kappa_3x_1x_3}{(\kappa_2+\kappa_3)\kappa_6}, \frac{\kappa_1\kappa_3(\kappa_5+\kappa_6)\kappa_7x_1^2x_3}{(\kappa_2+\kappa_3)\kappa_4\kappa_6(\kappa_8+\kappa_9)x_2}, \frac{\kappa_1\kappa_3(\kappa_5+\kappa_6)\kappa_7\kappa_9x_1^2x_3}{(\kappa_2+\kappa_3)\kappa_4\kappa_6(\kappa_8+\kappa_9)\kappa_{12}x_2} \Bigg). \end{aligned} \tag{4.2.10}$$

We rewrite the determinant of $M(\boldsymbol{x})$ using the positive parameterization given by $\Phi(x_1, x_2, x_3)$ in (4.2.10), and denote it by $p(x_1, x_2, x_3) = \det(M(\Phi(x_1, x_2, x_3)))$. In the light of Theorem 4.1.12, we are interested in the sign of $p(x_1, x_2, x_3)$ to study the multistationarity of the system.

The nonnegative solutions of the steady state equations determine the nonnegative steady states within the corresponding stoichiometric compatibility class. This system has at least one positive solution for any choice of parameters, but it can have up to three. The notion of multistationarity has already been defined in Section 4.1.2, and it

can be interpreted for the case of 2-site phosphorylation as in the following remark.

Remark 4.2.1. A vector of reaction rate constants $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_{12}) \in \mathbb{R}^{12}_{>0}$ *enables* multistationarity if there exist $\boldsymbol{c} = (c_1, c_2, c_3)$ such that the steady state system has at least two positive solutions, that is, with all coordinates positive. In this case we say that the network is multistationary in the linear invariant subspace with total amounts c_1, c_2, c_3. The vector κ is said to *preclude* multistationarity, if it does not enable it. ⬡

Remark 4.2.2. We note that there are no boundary steady state in any stoichiometric compatibility class $\mathscr{P}_{\boldsymbol{c}}$ such that $\mathscr{P}_{\boldsymbol{c}} \neq \emptyset$. In fact, the same holds for n-site for any $n \geq 2$ due to [CFMW17, Corollary 3]. ⬡

In [CM14], see also [CFMW17], sufficient conditions on the reaction rate constants for enabling or precluding multistationarity were given. These conditions arise from utilizing Theorem 4.1.12 with considering the Michaelis-Menten constants of each phosphorylation/dephosphorylation event. Michaelis-Menten constants arise from the enzyme catalyzed reactions such as phosphorylation or dephosphorylation, see e.g.[Lai78, Chapter 10] for detailed information on Michaelis-Menten constants. The Michaelis-Menten constants for 2-site phosphorylation cycle are given in [CFMW17, Page 40] as follows:

$$K_1 = \tfrac{\kappa_2+\kappa_3}{\kappa_1}, \quad K_2 = \tfrac{\kappa_5+\kappa_6}{\kappa_4}, \quad K_3 = \tfrac{\kappa_8+\kappa_9}{\kappa_7}, \quad K_4 = \tfrac{\kappa_{11}+\kappa_{12}}{\kappa_{10}}.$$

In order to employ the Michaelis-Menten constants, we define the following map

$$\begin{aligned} \pi \colon \mathbb{R}^{12}_{>0} &\to \mathbb{R}^{8}_{>0} \\ \boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_{12}) &\mapsto \boldsymbol{\eta} := (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}), \end{aligned} \tag{4.2.11}$$

and we further note that this is a continuous and surjective map. Under the map π, the positive parameterization Φ becomes

$$\Phi_{\boldsymbol{\eta}}(x_1, x_2, x_3) = \left(x_1, x_2, x_3, \frac{K_2\kappa_3 x_1 x_3}{K_1\kappa_6 x_2}, \frac{K_2K_4\kappa_3\kappa_9 x_1^2 x_3}{K_1K_3\kappa_6\kappa_{12}x_2^2}, \frac{x_1x_3}{K_1}, \frac{\kappa_3 x_1 x_3}{K_1\kappa_6}, \frac{K_2\kappa_3 x_1^2 x_3}{K_1K_3\kappa_6 x_2}, \frac{K_2K_3\kappa_3\kappa_9 x_1^2 x_3}{K_1\kappa_6\kappa_{12}x_2} \right).$$

After the positive parameterization given by the map $\Phi_{\boldsymbol{\eta}}$, the numerator of the determinant described in Theorem 4.1.12 becomes

$$
\begin{aligned}
p_{\boldsymbol{\eta}}(x) = {} & K_2\kappa_3(\kappa_3\kappa_{12}-\kappa_6\kappa_9)\Big(K_2K_4\kappa_3\kappa_9x_1^4x_3^2 + K_1K_3\kappa_6\kappa_{12}(x_1^3x_2^2x_3 + x_1^2x_2^3x_3 + x_1^2x_2^2x_3^2) \\
& + K_2K_3\kappa_3\kappa_{12}x_1^3x_2x_3^2\Big) + K_1K_2K_3\kappa_3\kappa_6\kappa_{12}((K_2+K_3)\kappa_3\kappa_{12} - (K_1+K_4)\kappa_6\kappa_9)x_1^2x_2^2x_3 \\
& + K_1\kappa_6\Big(K_2^2K_4\kappa_3^2\kappa_9^2\,x_1^4x_3 + 2K_2K_3K_4\kappa_3^2\kappa_9\kappa_{12}\,x_1^3x_2x_3 + K_1K_2K_3\kappa_3\kappa_6\kappa_{12}(\kappa_9+\kappa_{12})x_1^2x_2^3 \\
& + K_1K_2K_3K_4\kappa_3\kappa_6\kappa_9\kappa_{12}\,x_1^2x_2^2 + K_1K_3^2\kappa_6\kappa_{12}^2(\kappa_3+\kappa_6)\,x_1x_2^4 + 2\,K_1K_2K_3\kappa_3\kappa_6\kappa_{12}^2\,x_1x_2^3x_3 \\
& + K_1K_2K_3^2\kappa_3\kappa_6\kappa_{12}^2\,x_1x_2^3 + K_1K_3^2\kappa_6^2\kappa_{12}^2\,x_2^4x_3 + K_1^2K_3^2\kappa_6^2\kappa_{12}^2\,x_2^4\Big).
\end{aligned}
\tag{4.2.12}
$$

Note that the coefficients of $p_{\boldsymbol{\eta}}(x)$ depend on $\boldsymbol{\eta}$, and the sign of the determinant only depends on the numerator, since the denominator is always positive if $\boldsymbol{\eta}$ and x_1, x_2, x_3 are positive.

Remark 4.2.3. The ODE system in (4.2.3) is invariant under the map $\sigma\colon \mathbb{R}^{12}_{>0}\times\mathbb{R}^{9}_{>0} \to \mathbb{R}^{12}_{>0}\times\mathbb{R}^{9}_{>0}$, which is defined as the following symmetry of parameters and variables:

$$
\begin{aligned}
(\kappa_1,\dots,\kappa_{12},) &\mapsto (\kappa_{10},\kappa_{11},\kappa_{12},\kappa_7,\kappa_8,\kappa_9,\kappa_4,\kappa_5,\kappa_6,\kappa_1,\kappa_2,\kappa_3)\\
(x_1,\dots,x_9) &\mapsto (x_2,x_1,x_5,x_4,x_3,x_9,x_8,x_7,x_6).
\end{aligned}
$$

The reason is that the reaction network (4.2.2) remains invariant after interchanging E with F, S_0 with S_2, the intermediate complexes accordingly, and relabeling the reactions as the map above indicates. Under this map, we have

$$
\sigma(\boldsymbol{\kappa}) = \sigma(K_1,K_2,K_3,K_4,\kappa_3,\kappa_6,\kappa_9,\kappa_{12}) = (K_4,K_3,K_2,K_1,\kappa_{12},\kappa_9,\kappa_6,\kappa_3).
$$

It follows that $\boldsymbol{\eta}$ enables multistationarity if and only if $\sigma(\boldsymbol{\eta})$ does. In particular, any relation on the parameters that guarantees or precludes multistationarity, gives rise to a new relation after applying σ to all parameters. In many cases though, the relations are already invariant by σ. ⬡

Proposition 4.2.4 ([CM14, CFMW17])**.** With $p_{\boldsymbol{\eta}}$ as in (4.2.12), it holds that:

(Mono) If $p_{\boldsymbol{\eta}}(x)$ is positive for all $x_1,x_2,x_3>0$, then any $\kappa\in\pi^{-1}(\boldsymbol{\eta})$ does not enable multistationarity, and there is exactly one positive steady state in each invariant linear subspace.

(Mult) If $p_{\boldsymbol{\eta}}(x)$ is negative for some $x_1,x_2,x_3>0$, then any $\kappa\in\pi^{-1}(\boldsymbol{\eta})$ enables multistationarity in the invariant linear subspace containing the point

$$
\Phi_{\boldsymbol{\eta}}(x_1,x_2,x_3) = \left(x_1,x_2,x_3,\frac{K_2\kappa_3x_1x_3}{K_1\kappa_6x_2},\frac{K_2K_4\kappa_3\kappa_9x_1^2x_3}{K_1K_3\kappa_6\kappa_{12}x_2^2},\frac{x_1x_3}{K_1},\frac{\kappa_3x_1x_3}{K_1\kappa_6},\frac{K_2\kappa_3x_1^2x_3}{K_1K_3\kappa_6x_2},\frac{K_2K_3\kappa_3\kappa_9x_1^2x_3}{K_1\kappa_6\kappa_{12}x_2}\right).
$$

Proof. Proposition 4.2.4 is a specific instance of the Theorem 4.1.12, which is used to identify multistationarity for networks satisfying three conditions, namely dissipativity, absence of boundary steady states, and existence of an algebraic parameterization of the steady states [CFMW17]. The rank of the reaction network in (4.2.2) is 6, and it is dissipative since it is conservative, see (4.2.7). We pointed out in Remark 4.2.2 that the 2-site phosphorylation cycle does not have any boundary steady states in any stoichiometric compatibility class $\mathscr{P}_{\boldsymbol{c}}$ such that $\mathscr{P}_{\boldsymbol{c}} \neq \emptyset$. Furthermore, (4.2.10) yields a positive algebraic parameterization of the steady states. Therefore, the proof follows from Theorem 4.1.12. □

Explicitly, the polynomial $p_{\boldsymbol{\eta}}$ equals $\det(J_{\varphi_c}(\Phi_{\boldsymbol{\eta}}(x_1, x_2, x_3))$, where $\varphi_c \colon \mathbb{R}^9 \to \mathbb{R}^9$ is the function with first three components being the left-hand side of the equations in (4.2.8), and last 6 components being the right-hand side of $\frac{dx_4}{dt}, \ldots, \frac{dx_9}{dt}$ in (4.2.3), and J_F denotes the corresponding Jacobian. The Brouwer degree of $p_{\boldsymbol{\eta}}$ at zero is 1, and this is used to derive conditions (Mono) and (Mult) in Proposition 4.2.4 (see [CFMW17]).

We now have the ingredients to restate the conditions on the reaction rate constants that enable or preclude multistationarity given in [CM14]. Recall the map π from (4.2.11) and let

$$a(\boldsymbol{\eta}) = \kappa_3\kappa_{12} - \kappa_6\kappa_9, \qquad b(\boldsymbol{\eta}) = (K_2 + K_3)\kappa_3\kappa_{12} - (K_1 + K_4)\kappa_6\kappa_9, \tag{4.2.13}$$

where $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$.

Remark 4.2.5. Observe that $a(\boldsymbol{\eta})$ only depends on $\kappa_3, \kappa_6, \kappa_9, \kappa_{12}$. By letting

$$\boldsymbol{K} = (K_1, K_2, K_3, K_4), \qquad \overline{\boldsymbol{\kappa}} = (\kappa_3, \kappa_6, \kappa_9, \kappa_{12}),$$

it will be convenient sometimes to write $a(\overline{\boldsymbol{\kappa}})$ instead of $a(\boldsymbol{\eta})$. ⬡

The authors point out in [CM14] the following two observations about the multistationarity of 2-site phosphorylation network:

1. if $a(\boldsymbol{\eta}) > 0$ and $b(\boldsymbol{\eta}) > 0$ for some $\boldsymbol{\eta}$, then $\boldsymbol{\eta}$ precludes multistationarity,
2. if $a(\boldsymbol{\eta}) < 0$ for some $\boldsymbol{\eta}$, then $\boldsymbol{\eta}$ enables multistationarity.

The coefficients of the polynomial $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ given in (4.2.12) in the variables x_1, x_2, x_3 are polynomials in the eight parameters $K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}$. Five of these coefficients are positive multiples of $a(\boldsymbol{\eta})$, one is a positive multiple of $b(\boldsymbol{\eta})$, and the rest of the coefficients are positive. Note that some of the exponents corresponding to the coefficients which are positive multiples of $a(\boldsymbol{\eta})$ are vertices of the $\mathrm{New}(p_{\boldsymbol{\eta}}(\boldsymbol{x}))$. The statement (1) above easily follows from Proposition 4.2.4. In order to see why the statement (2) holds, we have to recall a necessary condition for the nonnegativity of $p_{\boldsymbol{\eta}}(x_1, x_2, x_3)$.

Proposition 4.2.6 (Lemma on Page 365, [Rez78])**.** If $p_{\boldsymbol{\eta}}(x_1, x_2, x_3)$ is nonnegative, then any term of $p_{\boldsymbol{\eta}}(x_1, x_2, x_3)$ that correspond to a vertex of New $(p_{\boldsymbol{\eta}}(x_1, x_2, x_3))$ cannot take negative sign.

It follows Proposition 4.2.6 that $p_{\boldsymbol{\eta}}(x_1, x_2, x_3)$ takes negative values, because the sign of the coefficients corresponding to some of its vertices are negative. In Proposition 4.2.7, we give a generalization of Proposition 4.2.6 which will play a crucial role in our strategy to study the case $a(\boldsymbol{\eta}) \geq 0$ and $b(\boldsymbol{\eta}) < 0$ open in [CM14] which was left open by Conradi and Mincheva.

We would like to mention that, apart from the description we provided above, the conditions for the existence of three positive steady states involving the parameters $\kappa_1, \ldots, \kappa_{12}$ and some of the total amounts are given in [FW12, BDG20], see also [CF12]. However, especially for the case of 2-site phosphorylation, we focus on describing the multistationarity in terms of $a(\boldsymbol{\eta})$ and $b(\boldsymbol{\eta})$. More specifically, we address to the cases $a(\boldsymbol{\eta}) \geq 0$ and $b(\boldsymbol{\eta}) < 0$ in Section 4.3. Furthermore, we give necessary conditions and sufficient conditions for multistationarity to arise in this case, and give an explicit parameterization of the boundary between the region of monostationarity and multistationarity. In view of Proposition 4.2.4, in order to determine what reaction rate constants $\boldsymbol{\kappa}$ enable multistationarity, we need to study what signs $p_{\boldsymbol{\eta}}$ attains over $\mathbb{R}^3_{>0}$, as a function of $\boldsymbol{\eta}$. To this end, we study the relation between the coefficients of $p_{\boldsymbol{\eta}}$ and the signs the polynomial attains using various algebraic techniques which we review in Section 4.2.2.

4.2.2 Introduction of Algebraic and Geometric Tools

The key results we will be employing in order to study the relation between the coefficients of a polynomial and the signs the polynomial attains, build on a geometric object, namely the Newton polytope. Consider a multivariate polynomial $p(x_1, \ldots, x_n) = \sum_{\boldsymbol{\alpha} \in A_p} p_\alpha \boldsymbol{x}^{\boldsymbol{\alpha}}$ in $\mathbb{R}[x_1, \ldots, x_n]$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$. Recall that the Newton polytope New $(f) \subseteq \mathbb{R}^n$ associated with f is the convex hull of the support of f, i.e., New $(f) =$ conv (A_f). For $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, the *supporting hyperplane* of New (f) with outer normal vector $\boldsymbol{v}$ is the hyperplane given by

$$\mathscr{H}(\text{New}\,(f)\,, \boldsymbol{v}) := \{\boldsymbol{x} \in \mathbb{R}^n : \ \langle \boldsymbol{x}, \boldsymbol{v} \rangle = h(\text{New}\,(f)\,, \boldsymbol{v})\}$$

where $h(\text{New}\,(f)\,, \boldsymbol{v}) = \sup\{\langle \boldsymbol{v}, \boldsymbol{y} \rangle : \ \boldsymbol{y} \in \text{New}\,(f)\}$. A *(nontrivial) face* of New (f) is the intersection of New (f) with a supporting hyperplane $\mathscr{H}(\text{New}\,(f)\,, \boldsymbol{v})$ given by some $\boldsymbol{v} \in \mathbb{R}^n$, and the *dimension of the face* F is the affine dimension of New $(f) \cap \mathscr{H}(\text{New}\,(f)\,, \boldsymbol{v})$. Given a face F of New (f), we define the *restriction of f to the monomials supported on*

F as

$$p_F(\boldsymbol{x}) := \sum_{\boldsymbol{\alpha} \in F} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}.$$

Furthermore, the *outer normal cone* of $\mathrm{New}(f)$ at the face F, i.e., $\mathcal{N}(\mathrm{New}(f), F)$, is the cone that consists of all $\boldsymbol{v} \in \mathbb{R}^n$ such that $F \subset H(\mathrm{New}(f), \boldsymbol{v})$. In other words, if $\mathrm{New}(f)$ is a d-dimensional polytope, then $\mathcal{N}(\mathrm{New}(f), F)$ is the cone generated by the outer normal vectors of the supporting hyperplanes of all the $d-1$ dimensional faces of $\mathrm{New}(p)$ containing F. Recall that we denote the interior of $\mathcal{N}(\mathrm{New}(f), F)$ with $\mathrm{int}(\mathcal{N}(\mathrm{New}(f), F))$. For more details on the theory of polytopes, we refer the reader to [Zie95], [Sch11] and [GO04, Chapter 16].

The first main property of the Newton polytope is that any nonzero sign attained by $p_F(\boldsymbol{x})$ also is attained by $p(\boldsymbol{x})$. This fact is a folklore in real algebraic geometry, but in Proposition 4.2.7 we give a rigorous statement and a short proof of this fact.

Proposition 4.2.7. Let $p \in \mathbb{R}[x_1, \ldots, x_n]$. Given a nonempty face F of $\mathrm{New}(p)$, consider the restriction p_F of p to the monomials supported on F. For any $\boldsymbol{x} \in \mathbb{R}^n_{>0}$ such that $p_F(\boldsymbol{x}) \neq 0$, there exists $\boldsymbol{y} \in \mathbb{R}^n_{>0}$ such that

$$\mathrm{sign}(p(\boldsymbol{y})) = \mathrm{sign}(p_F(\boldsymbol{x})).$$

Note that, the case when F is a zero dimensional face of $\mathrm{New}(p)$ was covered in [Rez78, Lemma on Page 365].

Proof. We find explicit values of $\boldsymbol{y}$ where the sign of $p(\boldsymbol{y})$ agrees with the sign of $p_F(\boldsymbol{x})$ as follows. For $p(x_1, \ldots, x_n) = \sum_{\boldsymbol{\alpha} \in A_p} p_\alpha \boldsymbol{x}^{\boldsymbol{\alpha}} \in \mathbb{R}[x_1, \ldots, x_n]$, without loss of generality we assume that $\mathrm{New}(p)$ is n-dimensional, and consider a d-dimensional face F of $\mathrm{New}(p)$ for some $d \leq n$. Let $\mathcal{N}_F$ denote the outer normal cone of $\mathrm{New}(p)$ at the face F. Then, for any vector $\boldsymbol{v} = (v_1, \ldots, v_n)$ in the interior of $\mathcal{N}_F$ (relative to the affine subspace of dimension $n-d$ containing it), the inner product $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ for $\boldsymbol{w} \in \mathrm{New}(f)$ is maximized when $\boldsymbol{w}$ belongs to the face F [Zie95, Lemma 2.8]. Let c denote the value of this inner product. Hence, given $\boldsymbol{x} \in \mathbb{R}^n_{>0}$, we have

$$p(x_1 t^{v_1}, \ldots, x_n t^{v_n}) = \sum_{\boldsymbol{\alpha} \in A_p} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} t^{\boldsymbol{v} \cdot \boldsymbol{\alpha}} = p_F(\boldsymbol{x}) t^c + \text{lower degree terms in } t.$$

Hence, the sign of $p(x_1 t^{v_1}, \ldots, x_n t^{v_n})$ agrees with the sign of $p_F(\boldsymbol{x})$ for $t \in \mathbb{R}_{>0}$ large enough. □

Example 4.2.8. Consider the polynomial $p(x, y) = y - 4xy^3 + x^2y^4 + 8x^3y^4$. The Newton polytope $\mathrm{New}(p)$ is a quadrilateral in the plane, see left panel in Figure 4.3. As $(1, 3)$ is a vertex, $p(x, y)$ attains negative values over $\mathbb{R}^2_{>0}$ by Proposition 4.2.7. To find a point where

p is negative, consider the outer normal cone at the vertex $(1,3)$, which is generated by the outer normal vectors $\boldsymbol{v_1} := (-2,1)$ and $\boldsymbol{v_2} := (-1,1)$. The vector $\boldsymbol{u} = \boldsymbol{v_1} + \boldsymbol{v_2} = (-3,2)$ belongs to its interior. Evaluation of p at (t^{-3}, t^2) is $-4t^3 + 2t^2 + 8t^{-1}$, which is negative for t larger than ≈ 1.34. ⎔

Example 4.2.9. Consider the polynomial $p(x,y) = 1 + x^2y^4 + x^4y^2 - 3x^3y^3$. The Newton polytope $\mathrm{New}(p)$ of p is a triangle in the plane and all of its vertices are positive, see middle panel in Figure 4.3. The edge F joining $(2,4)$ and $(4,2)$ contains a negative point $(3,3)$. We have

$$p_F(x,y) = x^2y^4 + x^4y^2 - 3x^3y^3 = x^2y^2(y^2 + x^2 - 3xy),$$

which is negative for instance when $x = y = 1$. It follows from Proposition 4.2.7 that p also attains negative values in $\mathbb{R}^2_{>0}$. To find an instance, we consider the outer normal cone at F, which is generated by one outer normal vector $\boldsymbol{u} = (1,1)$. Evaluation of p at (t,t) is $1 - t^6$, which is clearly negative for all $t > 1$. ⎔

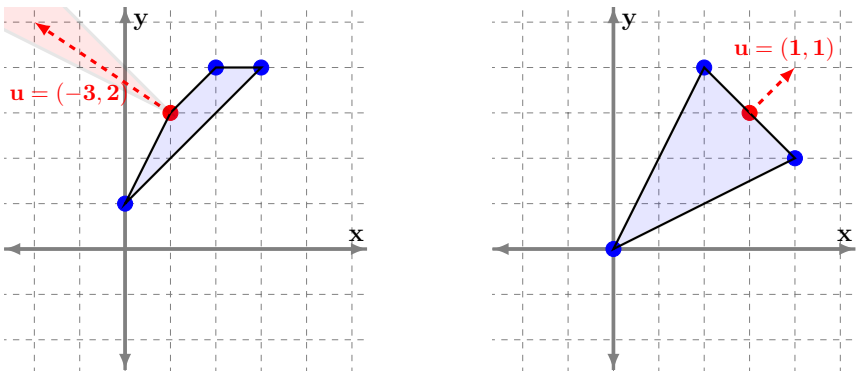

Figure 4.3: (Left) The quadrilateral corresponds is $\mathrm{New}(p)$ for p in Example 4.2.8, the shaded region is the outer normal cone at the vertex, and dashed vector is the chosen u. (Right) The triangle corresponds to the Newton polytope of p in Example 4.2.9, the dashed vector is the unique generator of the outer normal cone.

In what follows, we recall that a point $\boldsymbol{\alpha}$ in the support of a polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ is said to be *positive (negative)* if the coefficient of the monomial $x^{\boldsymbol{\alpha}}$ is positive (negative). Since we are primarily interested in the values of f over the positive orthant, the monomial $x^{\boldsymbol{\alpha}}$ is always evaluated positively. A useful consequence of Proposition 4.2.7 is the following result.

Corollary 4.2.10. Let $p \in \mathbb{R}[x_1, \ldots, x_n]$. Assume $\mathrm{New}(p)$ has dimension n and that all negative points A_p belongs to some proper face of $\mathrm{New}(p)$ (of dimension smaller than n). Then the following equivalence of statements holds:

$$p(\boldsymbol{x}) \geq 0 \text{ for all } \boldsymbol{x} \in \mathbb{R}^n_{>0} \qquad \text{if and only if} \qquad p(\boldsymbol{x}) > 0 \text{ for all } x \in \mathbb{R}^n_{>0}.$$

Proof. The reverse implication is clear. To prove the forward implication, decompose $p(\boldsymbol{x})$ as

$$p(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{\alpha} \in \partial(\mathrm{New}(p)) \\ p_{\boldsymbol{\alpha}} \neq 0}} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} + \sum_{\substack{\boldsymbol{\alpha} \in \mathrm{int}(\mathrm{New}(p)) \\ p_{\boldsymbol{\alpha}} \neq 0}} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}},$$

where $\partial(\mathrm{New}\,(p))$ and $\mathrm{int}\,(\mathrm{New}\,(p))$ denote the boundary and the interior of $\mathrm{New}\,(p)$ respectively. By assumption, the second summand has only positive coefficients and hence is positive over $\mathbb{R}^n_{>0}$. If $p(\boldsymbol{x}) = 0$ for some $\boldsymbol{x} \in \mathbb{R}^n_{>0}$, then necessarily the first summand is negative at this point $\boldsymbol{x}$, and it follows that the restriction of p to some proper face attains negative values. By Proposition 4.2.7, the same holds for p, contradicting that $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n_{>0}$. □

Our primary strategy for verifying multistationarity and monostationarity is to invoke Proposition 4.2.4, which requires us to symbolically certify the nonnegativity of a polynomial on the positive orthant. The theory of circuit polynomials, which we introduced in Chapter 2, will be our main tool for nonnegativity certification. Recall that a circuit polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ is a polynomial where support of f is an r-dimensional circuit for some $r \leq n$, see Definition 2.4.1. Furthermore, each circuit polynomial has an associated circuit number as defined in Definition 2.4.2, which can be used to check the nonnegativity of a circuit polynomial efficiently via Theorem 2.4.3. In Definition 2.4.1 the circuit polynomials are necessarily supported on circuits with even vertices, since otherwise they cannot be nonnegative due to [Rez78, Lemma on Page 365]. In 4, with an abuse of notation, we will use the name circuit polynomial to denote polynomials of the form

$$p(x) := c_\beta x^\beta + \sum_{j=0}^{r} c_{\alpha(j)} x^{\alpha(j)} \tag{4.2.14}$$

with $r \leq n$, coefficients $c_{\alpha(j)} \in \mathbb{R}_{>0}$, $c_\beta \in \mathbb{R}$, and exponents $\alpha(j), \beta \in \mathbb{N}^n$ such that $\mathrm{New}\,(p)$ is a simplex with vertices $\alpha(0), \ldots, \alpha(r)$ containing β in its interior. We note that these two definitions coincide when $\boldsymbol{x}$ is restricted to the positive orthant, since one can consider $p(x_1, \ldots, x_n) = q(x_1^2, \ldots, x_n^2)$; for further details see e.g., the discussion in [IdW16a, Section 3.1]. Moreover, the definition of the circuit number naturally extends to the case when vertices of the circuit are in $\mathbb{Z}^n$. With these considerations, the corollary that follows is a straightforward consequence of [IdW16a, Theorem 3.8]. It gives a way to check the nonnegativity of a circuit polynomial p over $\mathbb{R}^n_{>0}$ using the circuit number Θ_p.

Corollary 4.2.11 (Theorem 2.4.3)**.** A circuit polynomial p given as in (4.2.14) is non-negative over $\mathbb{R}^n_{\geq 0}$ if and only if

$$-p_{\boldsymbol{\beta}} \leq \Theta_p.$$

Now we summarize some general facts about univariate polynomials, which are repeatedly used in Section 4.3. Given a univariate polynomial with real coefficients $f = f_0 + f_1 x + \cdots + f_n x^n \in \mathbb{R}[x]$, the *discriminant of f* is a polynomial in $f_0, \ldots, f_n$, divides the parameter space $\mathbb{R}^{n+1}$ into regions where the number of *real roots* of $f(x)$ is constant. That is, in each connected component of the complement of Δ_f, the number of real roots of f is constant. Furthermore, if f has a multiple root, then $\Delta_f(f_0, \ldots, f_n) = 0$ and we say that the point $(f_0, \ldots, f_n)$ lies on the discriminant. Another practical fact that we will employ repeatedly to study the signs of an univariate polynomial is the result known as the *Descartes' rule of signs* in the literature. This classical result can be found in the elementary sources of algebra, see for example [BCR98, Chapter 1.2] or [Mes82, Chapter 4.11]. Furthermore, we refer to [Wan04] for a short and elegant proof of this fact.

Theorem 4.2.12 (Descartes Rule of Signs)**.** Given a univariate polynomial with real coefficients $f = f_0 + f_1 x + \cdots + f_n x^n \in \mathbb{R}[x]$, let $m \in \mathbb{N}$ be the number of sign changes in the sequence of coefficients $f_0, \ldots, f_n$ after removing the coefficients that are equal to zero. Then, the number positive roots (with multiplicity) of $f(x)$ equals to $m - 2k$ for some $k \in \mathbb{N}$.

Remark 4.2.13. By applying Theorem 4.2.12 to $g(x) = f(-x)$, a similar result can be achieved for the negative roots of $f(x)$. Hence, if m denotes the number of sign changes in the coefficient sequence that arise from $f(-x)$, then the number of negative roots of $f(x)$ is equal to $m - 2k$ for some $k \in \mathbb{N}$. ⬡

We further note Theorem 4.2.12 also holds when the exponents of x in $f(x)$ are real numbers, see [Wan04]. However, we use the Descartes' rule of signs in the context of polynomials exclusively.

Remark 4.2.14. In Section 4.3 we occasionally encounter homogeneous polynomials. Recall that a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ is homogeneous if the total degree of all of its monomials is the same, say d. In this case, $f(\lambda \boldsymbol{x}) = \lambda^d f(\boldsymbol{x})$ for any $\lambda \in \mathbb{R}$. Hence, the set of signs which f attains over $\mathbb{R}^n_{>0}$ agrees with the set of signs the polynomial $f(\lambda \boldsymbol{x})$ attains over $\mathbb{R}^n_{>0}$ for any choice of $\lambda > 0$. In particular, we can set one of the variables to 1, and study the signs of the resulting polynomial in the remaining $n - 1$ variables. ⬡

In Section 4.2.1, we reduced the problem of detecting multistationarity to studying the sign of the polynomial $p_{\boldsymbol{\eta}}(x_1, x_2, x_3)$ that is given in (4.2.12). We are particularly interested the monomial whose coefficient is multiple of $b(\boldsymbol{\eta})$, namely $x_1^2 x_2^2 x_3$, with exponent vector

$$\boldsymbol{m} := (2, 2, 1).$$

The Newton polytope of $p_{\boldsymbol{\eta}}$ depends on whether $a(\boldsymbol{\eta})$ vanishes or not. If $a(\boldsymbol{\eta}) \neq 0$, then $\mathrm{New}\,(p_{\boldsymbol{\eta}})$ is depicted in the left and middle panels of Figure 4.4 and has 10 vertices:

$$\mathrm{Vert}\,(\mathrm{New}\,(p_{\boldsymbol{\eta}})) = \big\{(4,0,2),(2,2,2),(4,0,1),(3,2,1),(2,3,1),(0,4,1),(2,3,0),(2,2,0),(1,4,0),(0,4,0)\big\}.$$

The point $\boldsymbol{m} = (2,2,1)$ is in the relative interior of the hexagonal face of $\mathrm{New}\,(p_{\boldsymbol{\eta}})$ depicted in the middle panel of Figure 4.4. The monomials with coefficient multiple of $a(\boldsymbol{\eta})$ are supported on the boundary of $\mathrm{New}\,(p_{\boldsymbol{\eta}})$. For $a(\boldsymbol{\eta}) = 0$, the corresponding Newton polytope is shown on the right panel of Figure 4.4. In this case $\boldsymbol{m}$ is an interior point of an edge of $\mathrm{New}\,(p_{\boldsymbol{\eta}})$. All other monomials have positive coefficients. The vertices of this Newton polytope are

$$(4,0,1),(2,3,0),(2,2,0),(1,4,0),(0,4,1),(0,4,0).$$

Let H be the face of $\mathrm{New}\,(p_{\boldsymbol{\eta}})$ containing $\boldsymbol{m}$: H is a hexagonal 2-dimensional face of $\mathrm{New}\,(p_{\boldsymbol{\eta}})$ if $a(\boldsymbol{\eta}) \neq 0$, and a 1-dimensional face if $a(\boldsymbol{\eta}) = 0$. Let $p_{\boldsymbol{\eta},H}$ denote the restriction of the polynomial $p_{\boldsymbol{\eta}}$ the face H of $\mathrm{New}\,(p_{\boldsymbol{\eta}})$.

Proposition 4.2.15. Let $p_{\boldsymbol{\eta}}$ be the polynomial given in (4.2.12), and let $a(\boldsymbol{\eta}), b(\boldsymbol{\eta})$ as in (4.2.13).

(i) $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ is either positive for all $\boldsymbol{x} \in \mathbb{R}^3_{>0}$ or attains negative values over $\mathbb{R}^3_{>0}$. Hence, κ enables multistationarity if and only if $p_{\boldsymbol{\eta}}$ attains negative values in $\mathbb{R}^3_{>0}$, where $\boldsymbol{\eta} = \pi(\boldsymbol{\kappa})$.

(ii) Assume $a(\boldsymbol{\eta}) \geq 0$. Then $\boldsymbol{\kappa}$ enables multistationarity if and only if $p_{\pi(\boldsymbol{\kappa}),H}$ attains negative values over $\mathbb{R}^3_{>0}$.

(iii) If $a(\boldsymbol{\eta}) \geq 0$ and $b(\boldsymbol{\eta}) \geq 0$, then any $\boldsymbol{\kappa} \in \pi^{-1}(\boldsymbol{\eta})$ precludes multistationarity and there is one positive steady state in each invariant linear subspace defined by the equations (4.2.8).

(iv) If $a(\boldsymbol{\eta}) < 0$, then any $\boldsymbol{\kappa} \in \pi^{-1}(\boldsymbol{\eta})$ enables multistationarity.

Proof. (i) Follows from Corollary 4.2.10 as coefficients of monomials supported on the interior of $\mathrm{New}\,(p_{\boldsymbol{\eta}})$ are positive; (ii) Follows from (i) and Proposition 4.2.7, as only $\boldsymbol{m} \in H$ can be a negative point; (iii) As $p_{\boldsymbol{\eta}}$ has only positive coefficients, the statement follows from (Mono) in Proposition 4.2.4; (iv) In this case four of the vertices are negative. From Proposition 4.2.7 we conclude that (Mult) in Proposition 4.2.4 holds. □

Statements (iii) and (iv) in Proposition 4.2.15 cover the two known cases from [CM14]. As $\boldsymbol{m}$ is not a vertex, $b(\boldsymbol{\eta}) < 0$ does not immediately guarantee that multistationarity

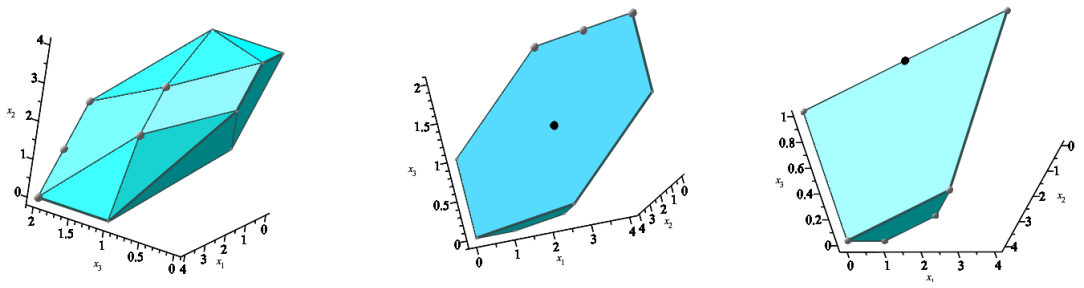

Figure 4.4: (Left and Middle) Newton polytope of the polynomial $p_{\boldsymbol{\eta}}$ in (4.2.12) for $a(\boldsymbol{\eta}) \neq 0$. The gray circles correspond to the monomials whose coefficient is a multiple of $a(\eta)$, and the black point to the monomial with coefficient a multiple of $b(\boldsymbol{\eta})$. (Right) Newton polytope of $p_{\boldsymbol{\eta}}$ when $a(\boldsymbol{\eta}) = 0$. The black point has coefficient a multiple of $b(\boldsymbol{\eta})$.

is enabled. In view of Proposition 4.2.15 (i), it only depends on $\pi(\boldsymbol{\kappa})$ whether $\boldsymbol{\kappa}$ enables multistationarity or not. Hence, we say that $\boldsymbol{\eta} \in \mathbb{R}^8_{>0}$ enables multistationarity if this is the case for any $\boldsymbol{\kappa} \in \pi^{-1}(\boldsymbol{\eta})$, or equivalently, if $p_{\boldsymbol{\eta}}$ attains negative values over $\mathbb{R}^3_{>0}$.

Corollary 4.2.16. The set $X \subseteq \mathbb{R}^8_{>0}$ of parameters $\boldsymbol{\eta}$ that enable multistationarity is open with the Euclidean topology in $\mathbb{R}^8_{>0}$.

Proof. By Proposition 4.2.15(i), $\boldsymbol{\eta} \in X$ if and only if $p_{\boldsymbol{\eta}}(x^*) < 0$ for some $\boldsymbol{x^*} \in \mathbb{R}^3_{>0}$. As $p_{\boldsymbol{\eta}}$ is continuous in the coefficients, there exists an open ball centered at $\boldsymbol{\eta}$ for which $p_{\boldsymbol{\eta'}}(\boldsymbol{x^*}) < 0$ for any $\boldsymbol{\eta'}$ in the ball. Hence X is open. □

Example 4.2.17. Consider $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}) = (343, 1, 1, 1, 2, 1, 1, 1)$, for which $a(\boldsymbol{\eta}) > 0$ and $b(\boldsymbol{\eta}) < 0$. Then, by Proposition 4.2.15 (ii), the parameter $\boldsymbol{\kappa} = (1, 341, 2, 2, 1, 1, 2, 1, 1, 2, 1, 1)$ enables multistationarity since $\pi^{-1}(\boldsymbol{\eta})$ and $p_{\boldsymbol{\eta},H}(7, 1, 49) = -24706290 < 0$.

In order to find a linear invariant subspace with multiple steady states, we use Proposition 4.2.7 to find a point where $p_{\boldsymbol{\eta}}(\boldsymbol{x}) < 0$. To this end, we note that $(-1, -1, 0)$ is an outer normal vector to H and consider

$$p_{\boldsymbol{\eta}}(7t^{-1}, t^{-1}, 49) = -\tfrac{24706290}{t^4} + \tfrac{38706521}{t^5}.$$

This expression is negative provided $t > \frac{47}{30}$. For example if $t = 2$, then $p_{\boldsymbol{\eta}}(7t^{-1}, t^{-1}, 49) = -\frac{10706059}{32} < 0$. Hence, the steady state defined by $(x_1, x_2, x_3) = (\frac{7}{2}, \frac{1}{2}, 49)$ satisfies (Mult) in Proposition 4.2.4. This steady state is $\boldsymbol{x^*} = \varphi(\frac{7}{2}, \frac{1}{2}, 49) = (\frac{7}{2}, \frac{1}{2}, 49, 2, 14, \frac{1}{2}, 1, 7, 7)$ and belongs to the linear invariant subspace defined by $E_{\text{tot}} = 11$, $F_{\text{tot}} = \frac{17}{2}$, $S_{\text{tot}} = \frac{161}{2}$. We solve the equations for the positive steady states in this linear invariant subspace, and obtain $\boldsymbol{x^*}$ together with two other positive steady states, given approximately by:

$(4.11, 0.91, 57.73, 1.51, 6.78, 0.7, 1.38, 6.2, 6.2)$, and $(3.43, 0.46, 47.21, 2.07, 15.6, 0.47, 0.94, 7.1, 7.1)$.

There are two other solutions with negative components. We see later in Example 4.3.12, how the initial parameter value and the point $(7, 1, 49)$ were chosen. ⬡

In Section 4.3, we study the open scenario $a(\boldsymbol{\eta}) \geq 0$ and $b(\boldsymbol{\eta}) < 0$ by focusing on $p_{\boldsymbol{\eta},H}$, see Proposition 4.2.15 (ii). We start by considering two strategies to certify that $p_{\boldsymbol{\eta},H}(x) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^3_{>0}$, which implies that multistationarity is precluded. Afterwards, we show that the polynomial $p_{\boldsymbol{\eta},H}(\boldsymbol{x})$ attains negative values for some $\boldsymbol{\eta}$, and finally, we provide an explicit parameterization of the boundary between the region in the parameter space where multistationarity is enabled and the region where it is precluded. In particular, given any vector of parameters, this gives means to certify whether multistationarity is enabled.

4.3 Results of the Case Study on 2-site Phosphorylation

We assume in this section that $a(\boldsymbol{\eta}) \geq 0$ and $b(\boldsymbol{\eta}) < 0$ and recall the face H of $\mathrm{New}\,(p_{\boldsymbol{\eta}})$ defined in Section 4.2.1. By Proposition 4.2.15(ii), $\boldsymbol{\eta}$ enables multistationarity if and only if $p_{\boldsymbol{\eta},H}$ attains negative values over $\mathbb{R}^3_{>0}$. In this section, we utilize various algebraic techniques to find conditions which imply the nonnegativity $p_{\boldsymbol{\eta},H}$. The face H belongs to the hyperplane $x_1 + x_2 = 4$, and hence $p_{\boldsymbol{\eta},H}$ is homogeneous of degree 4 in x_1, x_2. Therefore, by Remark 4.2.14, it suffices to study the signs of $p_{\boldsymbol{\eta},H}$ after setting $x_2 = 1$. By abuse of notation, we denote the restricted polynomial by $p_{\boldsymbol{\eta},H}(x_1, x_3)$. When $a(\boldsymbol{\eta}) \neq 0$, we have

$$\begin{aligned} p_{\boldsymbol{\eta},H}(x_1, x_3) = K_2\kappa_3 a(\boldsymbol{\eta})\Big(K_2K_4\kappa_3\kappa_9 x_1^4x_3^2 + K_2K_3\kappa_3\kappa_{12}x_1^3x_3^2 + K_1K_3\kappa_6\kappa_{12}x_1^2x_3^2\Big) \\ + K_1K_2K_3\kappa_3\kappa_6\kappa_{12}\, b(\boldsymbol{\eta})\, x_1^2x_3 + K_1\kappa_6\Big(K_2K_4\kappa_3\kappa_9\big(K_2\kappa_3\kappa_9\, x_1^4x_3 + 2K_3\kappa_3\kappa_{12}\, x_1^3x_3 \\ + K_1K_3\kappa_6\kappa_{12}\, x_1^2\big) + K_1K_3\kappa_6\kappa_{12}^2\big(K_1K_3\kappa_6 + 2\, K_2\kappa_3\, x_1x_3 + K_2K_3\kappa_3\, x_1 + K_3\kappa_6\, x_3\big)\Big). \end{aligned} \tag{4.3.1}$$

When $a(\boldsymbol{\eta}) = 0$, the polynomial of interest becomes:

$$p_{\boldsymbol{\eta},H}(x_1, x_3) = K_1\kappa_6\Big(K_2K_3\kappa_3^2\kappa_{12}^2((K_2+K_3)-(K_1+K_4))x_1^2x_3 + K_2K_4\kappa_3^2\kappa_9\big(K_2\kappa_9\, x_1^4x_3 + 2K_3\kappa_{12}\, x_1^3x_3\big) + K_1K_3\kappa_6\kappa_{12}^2\big(2\,K_2\kappa_3\, x_1x_3 + K_3\kappa_6\, x_3\big)\Big). \tag{4.3.2}$$

We distinguish between these two cases, mainly because the Newton polytope of $p_{\boldsymbol{\eta}}$ is different in each case, see left and right panels of Figure 4.4. We further note that the polynomial $p_{\boldsymbol{\eta},H}(x_1, x_3)$ is easier to deal with when $a(\boldsymbol{\eta}) = 0$, since some exponents of $p_{\boldsymbol{\eta},H}$ vanish in this case.

First, we present a sufficient condition for nonnegativity of $p_{\boldsymbol{\eta},H}$ which we obtain by studying the discriminant of a suitable polynomial in Section 4.3.1. Furthermore, this approach gives a complete characterization of the monostationarity region in the easier case of $a(\boldsymbol{\eta}) = 0$. Next, we consider an approach using circuit polynomials in Section 4.3.2, which gives an initial picture of the multistationarity region. In Section 4.3.3, with the help of initial picture, we give a parameterization of the boundary between multistationarity and monostationarity region. Lastly, in Section 4.3.4, we discuss the connectivity of multistationarity and monostationarity regions. Some proofs in Section 4.3.1 and Section 4.3.3 relies on symbolic computations done on `Maple`. We provide these computations on the supplementary file *SupplementaryInfoThesis.mw* in the end of the thesis, or alternatively found in the following link:

https://moto.math.nat.tu-bs.de/appliedalgebra_public/oguzhan_yuruk_thesis_supplementary_file

4.3.1 Necessary Polynomial Condition for Multistationarity via Discriminants

In this section we will find conditions that guarantee the nonnegativity of the polynomial of our interest, and utilize the discriminant and Descartes' rule of signs to make arguments about the positive roots of $p_{\boldsymbol{\eta},H}(x_1, x_3)$. A crucial observation in our strategy us that the polynomial $p_{\boldsymbol{\eta},H}(x_1, x_3)$ can be interpreted as a univariate polynomial in $\mathbb{R}[x_3]$. As a polynomial in $\mathbb{R}[x_3]$, the degree of $p_{\boldsymbol{\eta},H}$ is 2 if $a(\boldsymbol{\eta}) \neq 0$, and 1 $a(\boldsymbol{\eta}) = 0$. We study the coefficients of $p_{\boldsymbol{\eta},H} \in \mathbb{R}[x_3]$, which are parameterized by $\boldsymbol{\eta}\mathbb{R}^8_{\geq 0}$ and $x_1 \in \mathbb{R}_{>0}$. The study of the discriminant of $p_{\boldsymbol{\eta},H}$ leads to Theorem 4.3.1, whose proof relies on certain algorithms from the `Maple` package `RegularChains` based on [CDM+11]. For the proof of Theorem 4.3.1, we use the algorithms `RealRootCounting` which computes the number of distinct roots of a semi algebraic system, and `SamplePoints` which return at least one sample point per real connected component of a semi algebraic system.

Theorem 4.3.1. Let $\boldsymbol{\eta} \in \mathbb{R}^8_{>0}$ such that $a(\boldsymbol{\eta}) \geq 0$ and $b(\boldsymbol{\eta}) < 0$.

(i) Consider the following polynomial:

$$\begin{aligned} f(\boldsymbol{\eta}) :=& K_2^2K_3^2b(\boldsymbol{\eta})^4 - K_2K_3\kappa_3\kappa_{12}(K_1K_2^2 + K_3^2K_4)b(\eta)^3 + K_1K_2^2K_3^2K_4(\kappa_3^2\kappa_{12}^2 - 20\kappa_3\kappa_6\kappa_9\kappa_{12} - 8\kappa_6^2\kappa_9^2)b(\eta)^2 \\ &+ 18K_1K_2K_3K_4\kappa_3\kappa_6\kappa_9\kappa_{12}(\kappa_3\kappa_{12} + 2\kappa_6\kappa_9)(K_1K_2^2 + K_3^2K_4)b(\boldsymbol{\eta}) \\ &- K_1K_4\kappa_6\kappa_9\Big(27\kappa_3^2\kappa_6\kappa_9\kappa_{12}^2(K_1^2K_2^4 + K_3^4K_4^2) + 16K_1K_2^2K_3^2K_4(\kappa_3^3\kappa_{12}^3 - \kappa_6^3\kappa_9^3) \\ &+ 6K_1K_2^2K_3^2K_4\kappa_3\kappa_6\kappa_9\kappa_{12}(\kappa_3\kappa_{12} + 8\kappa_6\kappa_9)\Big). \end{aligned}$$

If $f(\boldsymbol{\eta}) \leq 0$, then $p_{\boldsymbol{\eta},H}$ is nonnegative over $\mathbb{R}^2_{>0}$, and $\boldsymbol{\eta}$ does not enable multistationarity.

(ii) Assume additionally that $a(\boldsymbol{\eta}) = 0$ and consider

$$g(\boldsymbol{K}) := K_2K_3(K_1 + K_4 - K_2 - K_3)^3 - 27K_1K_4(K_2 + K_3)(K_1K_2 - K_2K_3 + K_3K_4).$$

Then $p_{\boldsymbol{\eta},H}(x_1, x_3)$ is nonnegative over $\mathbb{R}^2_{>0}$ (and hence multistationarity is precluded) if and only if $g(\boldsymbol{K}) \leq 0$. Furthermore, $a(\boldsymbol{\eta}) = 0$ and $b(\boldsymbol{\eta}) < 0$ imply $K_1K_2 - K_2K_3 + K_3K_4 > 0$.

Proof. We observe that the coefficient of x_3 in $p_{\boldsymbol{\eta},H}$ in (4.3.1) and (4.3.2) is exactly $\kappa_6K_1q_{\boldsymbol{\eta}}(x_1)$ with

$$\begin{aligned} q_{\boldsymbol{\eta}}(x_1) :=& K_2^2K_4\kappa_3^2\kappa_9^2\,x_1^4 + 2K_2K_3K_4\kappa_3^2\kappa_9\kappa_{12}\,x_1^3 \\ &+ K_2K_3\,b(\eta)\,\kappa_3\kappa_{12}\,x_1^2 + 2K_1K_2K_3\kappa_3\kappa_6\kappa_{12}^2\,x_1 + K_1K_3^2\kappa_6^2\kappa_{12}^2. \end{aligned}$$

When $a(\boldsymbol{\eta}) = 0$, $p_{\boldsymbol{\eta},H}(x_1, x_3)$ is exactly $\kappa_6K_1q_\eta(x_1)x_3$ and it follows that $q_{\boldsymbol{\eta}}$ is nonnegative over $\mathbb{R}_{>0}$ if and only if $p_{\boldsymbol{\eta},H}$ is nonnegative over $\mathbb{R}^2_{>0}$. When $a(\boldsymbol{\eta}) > 0$, $p_{\boldsymbol{\eta},H}$ in (4.3.1) is a quadratic polynomial in x_3 with positive leading and constant terms. Therefore, if $q_{\boldsymbol{\eta}}$ is nonnegative over $\mathbb{R}_{>0}$, then $p_{\boldsymbol{\eta},H}$ is nonnegative over $\mathbb{R}^2_{>0}$.

Consequently, the theorem is proven if we show that: (1) Assuming $a(\boldsymbol{\eta}) \geq 0$ and $b(\boldsymbol{\eta}) < 0$, the polynomial $q_{\boldsymbol{\eta}}$ is nonnegative over $\mathbb{R}_{>0}$ if and only if $f(\boldsymbol{\eta}) \leq 0$, and (2) that this condition is equivalent to $g(\boldsymbol{K}) \leq 0$ when additionally $a(\boldsymbol{\eta}) = 0$.

We prove (1). The polynomial $q_{\boldsymbol{\eta}}$ has degree 4 in x_1, and only the coefficient of x_1^2 is negative (under the assumption $b(\boldsymbol{\eta}) < 0$). By Descartes' rule of signs, $q_{\boldsymbol{\eta}}$ has either two or zero positive roots and either two or zero negative roots (counted with multiplicity). Therefore, $q_{\boldsymbol{\eta}}$ attains negative values in $\mathbb{R}_{>0}$ if and only if $q_{\boldsymbol{\eta}}$ has two distinct positive roots.

Let Δ_{x_1} be the discriminant of $q_{\boldsymbol{\eta}}$; it is a polynomial in $\boldsymbol{\eta}$ and vanishes whenever $q_{\boldsymbol{\eta}}$ has a multiple root. We restrict the parameter space to the points where $b(\boldsymbol{\eta}) < 0$ and

$a(\boldsymbol{\eta}) \geq 0$ and define:

$$\Omega \;:=\; \{\boldsymbol{\eta} \in \mathbb{R}^8_{>0} \;:\; b(\boldsymbol{\eta}) < 0, a(\boldsymbol{\eta}) \geq 0 \text{ and } \Delta_{x_1}(\boldsymbol{\eta}) \neq 0\}.$$

In each connected component of Ω, the number of real roots of $q_{\boldsymbol{\eta}}$ is constant, and these are all simple roots. Since complex roots occur in pairs, the discriminant partitions $\mathbb{R}^8_{>0}$ into regions with four, two, or zero real roots. Now note that if $q_{\boldsymbol{\eta}}$ has four real roots, then necessarily two are positive and two are negative. Furthermore, in any connected component of Ω where $q_{\boldsymbol{\eta}}$ has two real roots, these are either both positive or both negative for all $\boldsymbol{\eta} \in \Omega$. This follows by continuity of the roots as a function of $\boldsymbol{\eta}$ in each connected component of Ω, together with the fact that $q_{\boldsymbol{\eta}}$ cannot have a positive and a negative root with multiplicity 1. We conclude that in every connected component of Ω, the number of positive real roots of $q_{\boldsymbol{\eta}}$ is also constant, and our goal is to determine the components where this number is 2.

We compute Δ_{x_1} and find that its zero set in Ω agrees with the zero set of one factor, f in the statement. Hence the sign of $f(\boldsymbol{\eta})$ in each connected component of Ω is constant. So the strategy to prove (1) is to show that $q_{\boldsymbol{\eta}}$ has two positive real roots if and only if $f(\boldsymbol{\eta}) > 0$, by checking that this is the case for at least one point in each connected component of Ω.

To select such points, we use the command `SamplePoints` from the `Maple` package `RegularChains`. To reduce the computational cost to effectively find the points, we make some simplifications. We note first that $b(\boldsymbol{\eta}), a(\boldsymbol{\eta})$ and $f(\boldsymbol{\eta})$ can be seen as polynomials in K_1, K_2, K_3, K_4 and the products $\kappa_3\kappa_{12}$ and $\kappa_6\kappa_9$, such that f is homogeneous of degree 8 in K_1, K_2, K_3, K_4 and homogeneous of degree 4 in $\kappa_3\kappa_{12}$ and $\kappa_6\kappa_9$; $a(\boldsymbol{\eta})$ and $b(\boldsymbol{\eta})$ are both homogeneous of degree 1 in $\kappa_3\kappa_{12}$ and $\kappa_6\kappa_9$; and $b(\boldsymbol{\eta})$ is homogeneous of degree 1 in K_1, K_2, K_3, K_4. Hence, given $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ and any $\lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0$, the point

$$\boldsymbol{\eta'} := \left(\lambda_1 K_1, \lambda_1 K_2, \lambda_1 K_3, \lambda_1 K_4, \tfrac{\lambda_2\lambda_3}{\lambda_4}\kappa_3, \lambda_2\kappa_6, \lambda_3\kappa_9, \lambda_4\kappa_{12}\right)$$

satisfies $f(\boldsymbol{\eta'}) = \lambda_1^8\lambda_2^4\lambda_3^4 f(\boldsymbol{\eta})$, $a(\boldsymbol{\eta'}) = \lambda_2\lambda_3 a(\boldsymbol{\eta})$ and $b(\boldsymbol{\eta'}) = \lambda_1\lambda_2\lambda_3 b(\boldsymbol{\eta})$. In particular, the signs of these three polynomials evaluated at $\boldsymbol{\eta}$ and $\boldsymbol{\eta'}$ agree, and $\boldsymbol{\eta}$ belongs to Ω, if and only if $\boldsymbol{\eta'}$ does, in which case both belong to the same connected component. As a consequence, it is enough to consider points of the form $(K_1, K_2, 1, K_4, \kappa_3, 1, 1, 1) \in \Omega$. The condition $a(\boldsymbol{\eta}) \geq 0$ becomes $\kappa_3 \geq 1$, and hence it is advantageous to reparameterize these points as $(K_1, K_2, 1, K_4, a+1, 1, 1, 1)$ with $a \geq 0$.

We have reduced the problem to selecting one point in each connected component of

$$\overline{\Omega} \;:=\; \{\boldsymbol{\eta} = \big(K_1, K_2, 1, K_4, a+1, 1, 1, 1\big) :\; K_1 > 0, K_2 > 0, K_4 > 0, a \geq 0, b(\boldsymbol{\eta}) < 0, f(\boldsymbol{\eta}) \neq 0\}.$$

To this end, we consider $f(\boldsymbol{\eta})$ for $\boldsymbol{\eta} \in \overline{\Omega}$ as a polynomial degree 4 in a in $\mathbb{R}[a]$ and coefficients in $\mathbb{R}[K_1, K_2, K_4]$, and we denote it as $\overline{f}_{\boldsymbol{v}}(a)$ for fixed $\boldsymbol{v} = (K_1, K_2, K_4) \in \mathbb{R}^3_{>0}$. We compute the discriminant Δ_a of $\overline{f}_{\boldsymbol{v}}$ with respect to a, which is a polynomial in K_1, K_2, K_4. The roots of the polynomial $\overline{f}_{\boldsymbol{v}}$ with variable a deform continuously in each connected component $C \subseteq \mathbb{R}^3_{>0}$ in the complement of $\Delta_a = 0$. Specifically, for a given point $\boldsymbol{v}$ in C, suppose $\overline{f}_{\boldsymbol{v}}$ has r real roots $\{a_1, \ldots, a_r\}$ for $r \leq 4$ such that $a_i \leq a_{i+1}$ for all i. For another point $\boldsymbol{v'}$ in C, $\overline{f}_{\boldsymbol{v'}}$ also has r roots $\{a'_1, \ldots, a'_r\}$ such that $a'_i \leq a'_{i+1}$ for all i. In C there exists a continuous path from $\boldsymbol{v}$ to $\boldsymbol{v'}$ such that a_i deforms continuously to a'_i. Therefore, there exists a continuous path in $\overline{\Omega}$ that takes a point from $\boldsymbol{v} \times (a_i, a_{i+1})$ to $\boldsymbol{v'} \times (a'_i, a'_{i+1})$.

Hence, in order to select at least one parameter point for each connected component of $\overline{\Omega}$, we consider first (at least) one choice of $\boldsymbol{v} = (K_1, K_2, K_4) \in \mathbb{R}^3_{>0}$ in each connected component C of the complement of $\Delta_a = 0$ with the command `SamplePoints`. We obtain a total of 22 points. For each of them, we find the nonnegative roots of $\overline{f}_{\boldsymbol{v}}$, i.e., f as a polynomial in a. Then, we extend K_1, K_2, K_4 to several parameter points in Ω' by selecting one value of a in each of the intervals the nonnegative roots define. This results in a list of points containing at least one point per connected component of Ω', and hence of Ω. Finally, for every such point $\boldsymbol{\eta}$, we find the number of positive roots of $q_{\boldsymbol{\eta}}$ (symbolically using the command `RealRootCounting` from the `Maple` package `RegularChains`) and determine the sign of $f(\boldsymbol{\eta})$. We conclude that $q_{\boldsymbol{\eta}}$ has two distinct positive real roots if and only if $f(\boldsymbol{\eta}) > 0$. It follows that $q_{\boldsymbol{\eta}}$ is nonnegative in $\mathbb{R}_{>0}$ if and only if $f(\boldsymbol{\eta}) \leq 0$, and in this case $p_{\boldsymbol{\eta},H}$ is nonnegative as well. This completes the proof of (1).

To prove (2), assume $a(\boldsymbol{\eta}) = 0$. It follows that $\kappa_3\kappa_{12} = \kappa_6\kappa_9$ and the condition $b(\boldsymbol{\eta}) < 0$ becomes $K_2 + K_3 < K_1 + K_4$. In this case,

$$f(\boldsymbol{\eta}) = \kappa_6^4\kappa_9^4(K_2 + K_3)(K_1K_2 - K_2K_3 + K_3K_4)g(\boldsymbol{K}).$$

Observe that under the assumption $b(\boldsymbol{\eta}) < 0$, we have

$$K_1K_2 + K_3K_4 > (K_1 + K_4) \cdot \min\{K_2, K_3\} > (K_2 + K_3) \cdot \min\{K_2, K_3\} > K_2K_3.$$

Hence $f(\boldsymbol{\eta}) > 0$ for $\boldsymbol{\eta} \in \Omega$ such that $a(\boldsymbol{\eta}) = 0$ if and only if $g(\boldsymbol{K}) > 0$. This concludes the proof. □

In the next example, we see how to apply Theorem 4.3.1 in the case $a(\boldsymbol{\eta}) = 0$ to characterize the regions of multistationarity and monostationarity. Note that in this case Theorem 4.3.1 gives a full description of the nonnegativity of $p_{\boldsymbol{\eta},H}$ in terms of the parameters K_1, K_2, K_3, K_4 .

Example 4.3.2. According to Theorem 4.3.1 (ii), if $a(\boldsymbol{\eta}) = 0$, then multistationarity is characterized by the inequality $g(\boldsymbol{K}) > 0$, which can be written as:

$$K_2K_3((K_1 + K_4) - (K_2 + K_3))^3 > 27K_1K_4(K_2 + K_3)(K_1K_2 + K_3K_4 - K_2K_3).$$

The expressions at each side of the inequality are positive when $b(\boldsymbol{\eta}) < 0$. A quick observation is that $g(\boldsymbol{K}) = 0$ intersects the two axes K_1 and K_4, because we have $g(K_2 + K_3, K_2, K_3, 0) = g(0, K_2, K_3, K_2 + K_3) = 0$. Fixing two of the four parameters in $\boldsymbol{K}$ gives a 2-dimensional slice of the zero set of the polynomial $g(\boldsymbol{K})$ and its complement. For example, if we fix $K_2 = K_3 = 1$, then the zero set of the polynomial

$$G(K_1, K_4) := g(K_1, 1, 1, K_4) = ((K_1 + K_4) - 2)^3K_1^3 - 54K_1K_4(K_1 + K_4 - 2))$$

describes a curve in (K_1, K_4)-plane that separates the regions of multistationarity and monostationarity, which was depicted in Figure 4.5. By checking whether $g(\boldsymbol{K})$ is positive or negative on points in the connected components of the complement of the given by $g(\boldsymbol{K}) = 0$, we can identify regions that correspond to monostationarity and multistationarity. A similar picture arise if we vary K_2 and K_3, a cartoon depiction of the regions of multistationarity and monostationarity for general $K_2, K_3 > 0$ illustrated in the right panel of Figure 4.5, see also Remark 4.3.3. ⎔

Remark 4.3.3. After setting $K_3 = 1$ as in the proof of Theorem 4.3.1, g becomes a polynomial in K_1, K_2 and K_4. The degree of g in K_1 and K_4 is 3. The discriminant of g with variables K_1 and K_4 is a polynomial in K_2, which does not vanish for any $K_2 > 0$. Therefore, for any $K_2 > 0$, the zero set of g in the (K_1, K_4)-plane is as depicted in the left panel of Figure 4.5. ⎔

Theorem 4.3.1 also allows us to work with the case $a(\boldsymbol{\eta}) > 0$. Unlike the previous case,

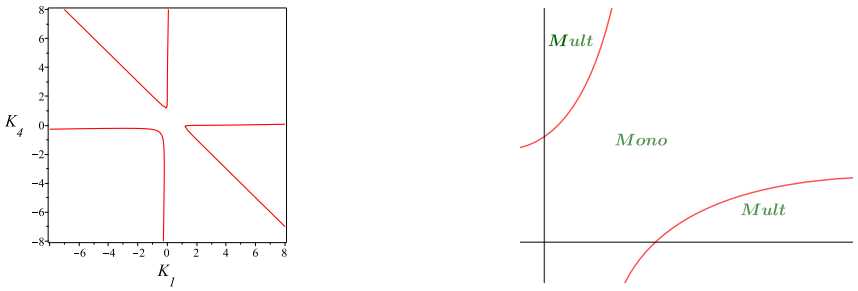

Figure 4.5: (Left) 2-dimensional section of the zero set of the polynomial g in Theorem 4.3.1. (Right) A nonrigorous sketch of the partition of the positive orthant into the regions of mono- and multistationarity.

we do not have the full description of the monostationarity region, because Theorem 4.3.1 only yields a sufficient condition for nonnegativity. Therefore, the region defined by $f(\boldsymbol{\eta}) > 0$ is contained in the monostationarity region, and using this fact one verify that a given $\boldsymbol{\eta}$ precludes multistationarity.

Example 4.3.4. Let $\boldsymbol{\eta} = (K_1, 1, 1, K_4, 2, 1, 1, 1)$ be a vector such that $K_1, K_4 > 0$ and note that $a(\boldsymbol{\eta}) > 0$ holds for any such $\boldsymbol{\eta}$. For this particular choice of $\boldsymbol{\eta}$ we have $b(\boldsymbol{\eta}) = 4 - (K_1 + K_4)$, and therefore, $b(\boldsymbol{\eta}) < 0$ if and only if $(K_1 + K_4) > 4$. Furthermore, the polynomial $f(\boldsymbol{\eta})$ in Theorem 4.3.1 reduces down to

$$\begin{aligned} F(K_1, K_4) := f(\boldsymbol{\eta}) = = 3K_1^4 - 284K_1^3K_4 - 590K_1^2K_4^2 - 284K_1K_4^3 + 3K_4^4 - 40K_1^3 + 808K_1^2K_4 \\ + 808K_1K_4^2 - 40K_4^3 + 192K_1^2 - 320K_1K_4 + 192K_4^2 - 384K_1 - 384K_4 + 256. \end{aligned}$$

The solution sets of $F(K_1, K_4) = 0$ and $b(\boldsymbol{\eta}) = 0$ in the (K_1, K_4)-plane are depicted as the red and the blue curves respectively in Figure 4.6. For example, consider $F(10, 1) = 5067 > 0$, Theorem 4.3.1 implies that $p_{\boldsymbol{\eta}}$ is nonnegative for all (K_1, K_4) in the gray(both dark and light) regions of the positive orthant. Moreover, the value of $b(\boldsymbol{\eta})$ is negative for this choice of K_1, K_4. This means that the dark grey region in Figure 4.6 the system is monostationary even though $b(\boldsymbol{\eta}) < 0$. We note this example points out an open subset that precludes multistationarity in the regions of the parameter space where $b(\boldsymbol{\eta}) < 0$, which was left open in the previous studies such as [CM14], [CFMW17]. ⬡

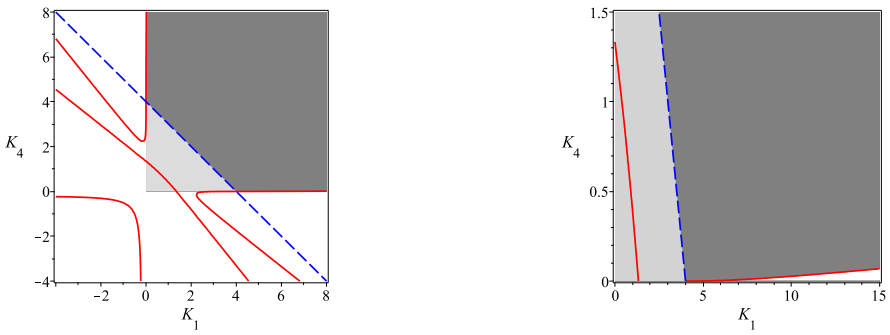

Figure 4.6: (Left) The solid-red curve is the solution set of $F(K_1, K_4) = 0$ from Example 4.3.4 in the (K_1, K_4)-plane, and the blue-dashed curve shows $b(\boldsymbol{\eta}) = 0$. In the gray region multistationarity is not enabled. The dark gray region is the one given in Theorem 4.3.1, where $F(K_1, K_4) < 0$ and $b(\boldsymbol{\eta}) < 0$. The light gray region shows $b(\boldsymbol{\eta}) \geq 0$ in the positive orthant. (Right) Zoom of the left panel for small K_4 and big K_1.

4.3.2 Necessary Condition for Multistationarity via Circuit Polynomials

Now we derive a necessary condition for multistationarity utilizing circuit polynomials. This new inequality, given in Theorem 4.3.5, allows for an easier inspection of the points verifying it, compared to Theorem 4.3.1 (see Corollary 4.3.10).

Since the case $a(\boldsymbol{\eta}) = 0$ is completely understood by Theorem 4.3.1, we focus on the case $a(\boldsymbol{\eta}) > 0$ and $b(\boldsymbol{\eta}) < 0$. Consider the Newton polytope H of $p_{\boldsymbol{\eta},H}(x_1, x_3)$ in (4.3.1) for $a(\boldsymbol{\eta}) \neq 0$. This polytope is the convex hull of $A_{p_{\boldsymbol{\eta},H}}$,i.e. the support of $p_{\boldsymbol{\eta},H}$ as a polynomial in $\mathbb{R}[x_1, x_3]$. We label the exponents in $A_{p_{\boldsymbol{\eta},H}}$ as follows (see left panel of Figure 4.7):

$$\begin{gathered}\boldsymbol{\alpha_1} := (4,2),\ \boldsymbol{\alpha_2} := (2,2),\ \boldsymbol{\alpha_3} := (0,1),\ \boldsymbol{\alpha_4} := (4,1),\ \boldsymbol{\alpha_5} := (2,0),\ \boldsymbol{\alpha_6} := (0,0),\\ \boldsymbol{m} := (2,1),\ \boldsymbol{b_1} := (3,2),\ \boldsymbol{b_2} := (1,0),\ \boldsymbol{i_1} := (3,1),\ \boldsymbol{i_2} := (1,1).\end{gathered} \tag{4.3.3}$$

Note that $A_{p_{\boldsymbol{\eta},H}}$ is very well structured: $\boldsymbol{m}$ is the barycenter of the two triangles given by the vertices $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_3}, \boldsymbol{\alpha_5}$ and $\boldsymbol{\alpha_2}, \boldsymbol{\alpha_4}, \boldsymbol{\alpha_6}$; $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ are the midpoints of the two edges of H given by $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_5}, \boldsymbol{\alpha_6}$ respectively; $\boldsymbol{i_1}$ and $\boldsymbol{i_2}$ are in the interior of H; and finally $\boldsymbol{m}$ is the midpoint of both $\boldsymbol{b_1}, \boldsymbol{b_2}$ and $\boldsymbol{i_1}, \boldsymbol{i_2}$. We exploit this structure to decompose $p_{\boldsymbol{\eta},H}(x_1, x_3)$ into the sum of four circuit polynomials with associated simplices with vertices $\{\boldsymbol{\alpha_1}, \boldsymbol{\alpha_3}, \boldsymbol{\alpha_5}\}$, $\{\boldsymbol{\alpha_2}, \boldsymbol{\alpha_4}, \boldsymbol{\alpha_6}\}$, $\{\boldsymbol{b_1}, \boldsymbol{b_2}\}$ and $\{\boldsymbol{i_1}, \boldsymbol{i_2}\}$.

Let $p_{\boldsymbol{\eta},1}$ be the circuit polynomial which is supported on the exponent $\boldsymbol{m}$ as inner term and 2-dimensional simplex $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_3}, \boldsymbol{\alpha_5}$, and given as follows:

$$p_{\boldsymbol{\eta},1}(x_1, x_3) = c_{\boldsymbol{\eta},\boldsymbol{\alpha_1}}\boldsymbol{x}^{\boldsymbol{\alpha_1}} + c_{\boldsymbol{\eta},\boldsymbol{\alpha_2}}\boldsymbol{x}^{\boldsymbol{\alpha_2}} + c_{\boldsymbol{\eta},\boldsymbol{\alpha_3}}\boldsymbol{x}^{\boldsymbol{\alpha_3}} + c_{\boldsymbol{\eta},\boldsymbol{m}}\boldsymbol{x}^{\boldsymbol{m}}$$

where $c_{\boldsymbol{\eta},\boldsymbol{\alpha_i}}$ and $c_{\boldsymbol{\eta},\boldsymbol{m}}$ are exactly the coefficients of $\boldsymbol{x}^{\boldsymbol{\alpha_i}}$ and $\boldsymbol{x}^{\boldsymbol{m}}$ in $p_{\boldsymbol{\eta},H}(\boldsymbol{x})$. Invoking Corollary 4.2.11 on $p_{\boldsymbol{\eta},1}$, yields that $p_{\boldsymbol{\eta},1}$ nonnegative if and only if $-c_{\boldsymbol{\eta},\boldsymbol{m}} \leq \Theta_{p_{\boldsymbol{\eta},1}}$, where $\Theta_{p_{\boldsymbol{\eta},1}}$ is the circuit number of $p_{\boldsymbol{\eta},1}$. The nonnegativity of $p_{\boldsymbol{\eta},1}$ is a sufficient condition for the nonnegativity of $p_{\boldsymbol{\eta}}$, and hence a sufficient condition for monostationarity in 2-site case. Consider another circuit polynomial $p_{\boldsymbol{\eta},2}$, that is supported on the exponent $\boldsymbol{m}$ as inner term and 2-dimensional simplex $\boldsymbol{\alpha_2}, \boldsymbol{\alpha_4}, \boldsymbol{\alpha_6}$, whose coefficients are defined in a similar manner to $p_{\boldsymbol{\eta},1}$. Corollary 4.2.11 yields another necessary condition, i.e., $-c_{\boldsymbol{\eta},\boldsymbol{m}} \leq \Theta_{p_{\boldsymbol{\eta},2}}$, for nonnegativity of $p_{\boldsymbol{\eta}}$. Therefore, we observe that if $-c_{\boldsymbol{\eta},\boldsymbol{m}} \leq \Theta_{p_{\boldsymbol{\eta},1}} + \Theta_{p_{\boldsymbol{\eta},2}}$, then $p_{\boldsymbol{\eta}}$ is nonnegative. In Theorem 4.3.5, we extend this approach by also considering circuit polynomials supported by the 1-dimensional simplices $\{\boldsymbol{b_1}, \boldsymbol{b_2}\}$ and $\{\boldsymbol{i_1}, \boldsymbol{i_2}\}$. We further note that, the method we incorporate in Theorem 4.3.5 is one of the pioneering applications of circuit polynomials, which we expect to be extendable to the cases of higher site phosphorylation or different networks.

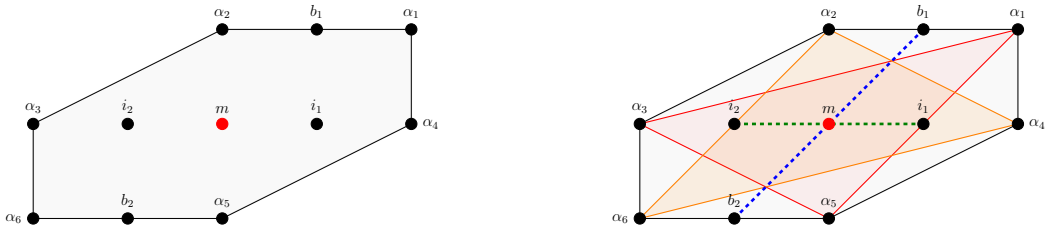

Figure 4.7: (Left) An illustration of H and support set $A_{p_{\boldsymbol{\eta},H}}$, where $\boldsymbol{\alpha_j}, \boldsymbol{b_j}, \boldsymbol{i_j}$ are as in (4.3.3). (Right) The circuits of the SONC decomposition, consisting of two 2-dimensional circuits with vertices $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_3}, \boldsymbol{\alpha_5}$ and $\boldsymbol{\alpha_2}, \boldsymbol{\alpha_4}, \boldsymbol{\alpha_6}$, and two 1-dimensional circuits, with vertices $\boldsymbol{b_1}, \boldsymbol{b_2}$ and $\boldsymbol{i_1}, \boldsymbol{i_2}$.

Theorem 4.3.5. Assume $a(\boldsymbol{\eta}) \geq 0$ and $b(\boldsymbol{\eta}) < 0$. If

$$-b(\boldsymbol{\eta}) \leq 3\big(K_1K_4\kappa_6^2\kappa_9^2a(\eta)\big)^{\frac{1}{3}}\Big(K_1^{\frac{1}{3}} + K_4^{\frac{1}{3}}\Big) + 4\left(K_1K_4\kappa_3\kappa_6\kappa_9\kappa_{12}\right)^{\frac{1}{2}} + 2\left(K_2K_3\kappa_3\kappa_{12}a(\boldsymbol{\eta})\right)^{\frac{1}{2}}, \tag{4.3.4}$$

then $p_{\boldsymbol{\eta},H}$ is nonnegative over $\mathbb{R}^2_{>0}$, and hence $\boldsymbol{\eta}$ does not enable multistationarity.

Proof. Assume $a(\boldsymbol{\eta}) > 0$. We write $p_{\boldsymbol{\eta},H}(x)$ as the sum of four circuit polynomials. Let $p_{\boldsymbol{\eta},1}$ be a circuit polynomial which has the exponent $\boldsymbol{m}$ as inner term and 2-dimensional simplex $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_3}, \boldsymbol{\alpha_5}$ as follows,

$$p_{\boldsymbol{\eta},1}(x_1, x_3) = c_{\boldsymbol{\eta},\boldsymbol{\alpha_1}}\boldsymbol{x}^{\boldsymbol{\alpha_1}} + c_{\boldsymbol{\eta},\boldsymbol{\alpha_2}}\boldsymbol{x}^{\boldsymbol{\alpha_2}} + c_{\boldsymbol{\eta},\boldsymbol{\alpha_3}}\boldsymbol{x}^{\boldsymbol{\alpha_3}} + \bar{c}_{\boldsymbol{\eta},1}\boldsymbol{x}^{\boldsymbol{m}}$$

where $c_{\boldsymbol{\eta},\boldsymbol{\alpha_i}}$ is exactly the coefficient of $\boldsymbol{x}^{\boldsymbol{\alpha_i}}$ in $p_{\boldsymbol{\eta},H}(\boldsymbol{x})$, and $\bar{c}_{\boldsymbol{\eta},1}$ is in $\mathbb{R}$. Similarly, we define the circuit polynomials $p_{\boldsymbol{\eta},2}, p_{\boldsymbol{\eta},3}, p_{\boldsymbol{\eta},4}$ with exponent $\boldsymbol{m}$ as inner term with 2-dimensional simplex $\boldsymbol{\alpha_2}, \boldsymbol{\alpha_4}, \boldsymbol{\alpha_6}$, and 1-dimensional simplices $\boldsymbol{b_1}, \boldsymbol{b_2}$ and $\boldsymbol{i_1}, \boldsymbol{i_2}$ respectively. As before, we let $\bar{c}_{\boldsymbol{\eta},i}$ be the coefficient of $\boldsymbol{x}^{\boldsymbol{m}}$ in the respective polynomial $p_{\boldsymbol{\eta},i}$. Furthermore, the coefficients of remaining terms in each $p_{\boldsymbol{\eta},i}$ is assumed to be equal to the coefficient of the same term in $p_{\boldsymbol{\eta},H}$. The Newton polytopes of these circuit polynomials are illustrated in the right panel of Figure 4.7.

The circuit number corresponding to each of the circuit polynomials are:

$$\Theta_{p_{\boldsymbol{\eta},1}} = 3(c_{\boldsymbol{\eta},\boldsymbol{\alpha_1}}c_{\boldsymbol{\eta},\boldsymbol{\alpha_3}}c_{\boldsymbol{\eta},\boldsymbol{\alpha_5}})^{\frac{1}{3}}, \qquad \Theta_{p_{\boldsymbol{\eta},2}} = 3(c_{\boldsymbol{\eta},\boldsymbol{\alpha_2}}c_{\boldsymbol{\eta},\boldsymbol{\alpha_4}}c_{\boldsymbol{\eta},\boldsymbol{\alpha_6}})^{\frac{1}{3}},$$
$$\Theta_{p_{\boldsymbol{\eta},3}} = 2(c_{\boldsymbol{\eta},\boldsymbol{b_1}}c_{\boldsymbol{\eta},\boldsymbol{b_2}})^{\frac{1}{2}}, \qquad \Theta_{p_{\boldsymbol{\eta},4}} = 2(c_{\boldsymbol{\eta},\boldsymbol{i_1}}c_{\boldsymbol{\eta},\boldsymbol{i_2}})^{\frac{1}{2}}.$$

Now assume that the following inequality is satisfied for $c_{\boldsymbol{\eta},m}$, the coefficient of x^m in $p_{\boldsymbol{\eta},H}$:

$$-c_{\boldsymbol{\eta},\boldsymbol{m}} \leq \Theta_{p_{\boldsymbol{\eta},1}} + \Theta_{p_{\boldsymbol{\eta},2}} + \Theta_{p_{\boldsymbol{\eta},3}} + \Theta_{p_{\boldsymbol{\eta},4}}. \tag{4.3.5}$$

Then one can find $\bar{c}_{\boldsymbol{\eta},1}, \bar{c}_{\boldsymbol{\eta},2}, \bar{c}_{\boldsymbol{\eta},3}, \bar{c}_{\boldsymbol{\eta},4} \in \mathbb{R}$ such that $\sum \bar{c}_{\boldsymbol{\eta},i} = c_{\boldsymbol{\eta},\boldsymbol{m}}$ and for all i, $-\bar{c}_{\boldsymbol{\eta},i} \leq \Theta_{p_{\boldsymbol{\eta},i}}$. Corollary 4.2.11 implies that each $p_{\boldsymbol{\eta},i}$ is nonnegative over $\mathbb{R}^2_{>0}$. As $p_{\boldsymbol{\eta},H} = p_{\boldsymbol{\eta},1} + p_{\boldsymbol{\eta},2} + p_{\boldsymbol{\eta},3} + p_{\boldsymbol{\eta},4}$, $p_{\boldsymbol{\eta},H}$ also is nonnegative. In terms of the entries of $\boldsymbol{\eta}$, (4.3.5) becomes

$$\begin{aligned}-K_1K_2K_3\kappa_3\kappa_6\kappa_{12}b(\boldsymbol{\eta}) \leq\ & 3K_1K_2K_3\kappa_3\kappa_6\kappa_{12}\left(K_1K_4^2\kappa_6^2\kappa_9^2a(\boldsymbol{\eta})\right)^{1/3} + 3\left(K_1^5K_2^3K_3^3K_4\kappa_3^3\kappa_6^5\kappa_9^2\kappa_{12}^3a(\boldsymbol{\eta})\right)^{1/3} \\ &+ 4\left(K_1^3K_2^2K_3^2K_4\kappa_3^3\kappa_6^3\kappa_9\kappa_{12}^3\right)^{1/2} + 2\left(K_1^2K_2^3K_3^3\kappa_3^3\kappa_6^2\kappa_{12}^3a(\boldsymbol{\eta})\right)^{1/2},\end{aligned}$$

which after factoring out terms and simplifying gives the inequality in the statement. □

Remark 4.3.6. The SONC decomposition of $p_{\boldsymbol{\eta}}$ into $p_{\boldsymbol{\eta},1}, p_{\boldsymbol{\eta},2}, p_{\boldsymbol{\eta},3}, p_{\boldsymbol{\eta},4}$ in the proof of Theorem 4.3.5 is not unique. Other sufficient conditions may be derived using other covers of H, see e.g., [DIdW19, page 20]. Two main reasons underlie the choice of this particular cover. First, it uses the least possible number of circuits while using every positive point only once. Hence we use all the possible positive weight and avoid introducing new parameters for nondisjoint circuits. Second, as $\boldsymbol{m}$ is the barycenter of each chosen circuit, the derived circuit numbers have simple expressions. ⬡

Remark 4.3.7. Due to [FdW19, Theorem 4.4], $p_{\boldsymbol{\eta}}$ is nonnegative if and only if $p_{\boldsymbol{\eta}} \in \mathscr{C}_{A_{p_{\boldsymbol{\eta}}}}$, see Definition 2.4.8. This means that, whenever $p_{\boldsymbol{\eta}}$ is nonnegative, there must be a decomposition of $p_{\boldsymbol{\eta}}$ into nonnegative circuit polynomials, which one can use to derive a condition such as (4.3.4). Therefore, the approach we used in the proof of Theorem 4.3.5 completely captures the nonnegativity of $p_{\boldsymbol{\eta}}$. ⬡

Example 4.3.8. To illustrate the use of inequality (4.3.4) to certify monostationarity, consider $\boldsymbol{\eta} = (2, 0.5, 0.5, 2, 2, 1, 1, 1)$. Then, (4.3.4) holds since the right hand side is ≈ 24.72, while the left hand side is 2. By Theorem 4.3.5, $\boldsymbol{\eta}$ does not enable multistationarity. Indeed, $p_{\boldsymbol{\eta},H}(x_1, x_3) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^2_{\geq 0}$, as it also can be seen by rewriting the polynomial as:

$$\begin{aligned}p_{\boldsymbol{\eta},\boldsymbol{H}}(x_1, x_3) = {} & x_2^4x_3 + 4x_1^4x_3 + \tfrac{1}{2}x_1^3x_2x_3^2 + 8x_1^3x_2x_3 + x_1^2x_2^2x_3^2 \\ & + 4x_1^2x_2^2 + 4x_1x_2^3x_3 + x_1x_2^3 + x_1^4x_3^2 + x_2^4 + (x_1^2x_3 - x_2^2)^2.\end{aligned}$$

⬡

Example 4.3.9. We fix the parameters $(K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}) = (1, 1, 2, 1, 1, 1)$ as in Example 4.3.4. Figure 4.8 shows a comparison of the two necessary conditions for multistationarity from Theorem 4.3.5 and Theorem 4.3.1. For this choice of parameters, inequality (4.3.4) becomes

$$0 \leq 3\left(K_1K_4^2\right)^{\frac{1}{3}} + 4\sqrt{2}\left(K_1K_4\right)^{\frac{1}{2}} + 3\left(K_1^2K_4\right)^{\frac{1}{3}} - K_1 - K_4 + 2\sqrt{2} + 4. \tag{4.3.6}$$

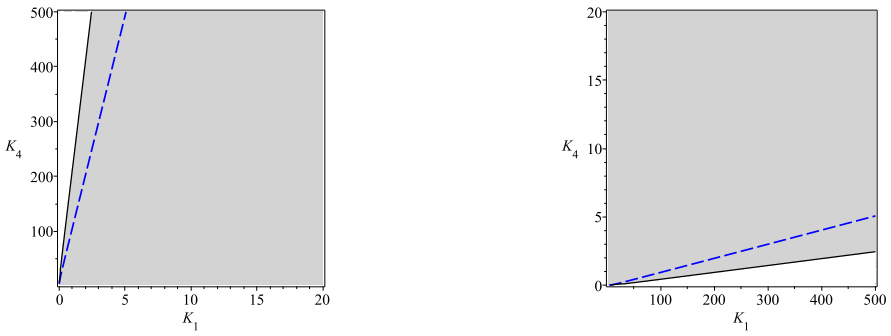

Figure 4.8: For $(K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}) = (1, 1, 2, 1, 1, 1)$, the region between blue dashed lines is the region where we can certify monostationarity using Theorem 4.3.1. The region given between full lines is the region where we can certify monostationarity using Theorem 4.3.5. The two panels focus on either K_1 large or K_1 small.

Figure 4.8 hints at that the sufficient condition for monostationarity of Theorem 4.3.5 includes a cone pointed at zero. To investigate this further, consider the line $sK_1 = K_4$ for $s \in (0, +\infty)$. Then the right hand side of (4.3.6) becomes

$$\left(3s^{\frac{2}{3}} + 4\sqrt{2}s^{\frac{1}{2}} + 3s^{\frac{1}{3}} - s - 1\right)K_1 + \left(2\sqrt{2} + 4\right). \tag{4.3.7}$$

The positive half ray belongs to the monostationarity region if (4.3.7) is positive for all $K_1 > 0$. As (4.3.7) is linear in K_1 with positive constant term, it is positive for all $K_1 > 0$ if and only if the leading coefficient is positive. This holds if and only if s lies in the interval $\approx (1/197.995, 197.995)$. ⬡

The conclusions in Example 4.3.9 extend to any choice of fixed parameters $K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}$. In particular, in the (K_1, K_4) plane, the region of monostationarity includes a cone pointed at zero that includes the line $K_1 = K_4$. This is the content of the next corollary. This result will be critical to obtain a parametric description of the regions of mono- and multistationarity in Section 4.3.3 (see Lemma 4.3.13).

Corollary 4.3.10. Assume that $\boldsymbol{\eta'} := (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ is fixed so that $a(\overline{\boldsymbol{\kappa}}) = a(k_3, k_6, k_9, k_{12}) \geq 0$ and consider the line $K_4 = sK_1$ in $\mathbb{R}^2_{>0}$ with coordinates K_1, K_4. There exist $0 < s_1(\boldsymbol{\eta'}) < \frac{14-\sqrt{192}}{2}$ and $\frac{14+\sqrt{192}}{2} < s_2(\boldsymbol{\eta'})$ such that:

(i) For any $s \in [s_1(\boldsymbol{\eta'}), s_2(\boldsymbol{\eta'})]$, the points in the line $K_4 = sK_1$ satisfy inequality (4.3.4).

(ii) If $s \notin [s_1(\boldsymbol{\eta'}), s_2(\boldsymbol{\eta'})]$, then there exists K_1' such that (4.3.4) holds if and only if $K_1 \leq K_1'$.

(iii) If $\kappa_3\kappa_{12}$ increases, while $K_2, K_3, \kappa_6, \kappa_9$ remain fixed, then $s_1(\boldsymbol{\eta'})$ decreases to zero and $s_2(\boldsymbol{\eta'})$ increases to $+\infty$.

In particular, if $K_1 = K_4$ and $a(\overline{\boldsymbol{\kappa}}) \geq 0$, multistationarity is not enabled.

Proof. As $\boldsymbol{\eta'}$ and $\overline{\boldsymbol{\kappa}}$ are fixed, inequality (4.3.4) is a relation on K_1 and K_4. We rewrite it as:

$$\begin{aligned} 0 \;\leq\; & -(K_1+K_4)\kappa_6\kappa_9 + 3\big(K_1K_4\kappa_6^2\kappa_9^2 a(\overline{\boldsymbol{\kappa}})\big)^{\frac{1}{3}}\Big(K_4^{\frac{1}{3}} + K_1^{\frac{1}{3}}\Big) \;+\; 4\,(K_1K_4\kappa_3\kappa_6\kappa_9\kappa_{12})^{\frac{1}{2}} \\ & + 2\,(K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}}))^{\frac{1}{2}} + (K_2+K_3)\kappa_3\kappa_{12}. \end{aligned}$$

When $K_4 = sK_1$, this inequality becomes

$$\begin{aligned} 0 \;\leq\; & \Big(-(1+s)\kappa_6\kappa_9 + 3(s\kappa_6^2\kappa_9^2 a(\overline{\boldsymbol{\kappa}}))^{\frac{1}{3}}(s^{\frac{1}{3}}+1) \;+\; 4(s\kappa_3\kappa_6\kappa_9\kappa_{12})^{\frac{1}{2}}\Big)K_1 \\ & +\; 2(K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}}))^{\frac{1}{2}} + (K_2+K_3)\kappa_3\kappa_{12}. \end{aligned} \tag{4.3.8}$$

First, note that since by assumption $\kappa_3\kappa_{12} \geq \kappa_6\kappa_9$, we have:

$$(1+s)\kappa_6\kappa_9 = (1+s)\big(\kappa_6^2\kappa_9^2\big)^{\frac{1}{2}} \leq (1+s)(\kappa_3\kappa_6\kappa_9\kappa_{12})^{\frac{1}{2}}.$$

Hence, if $(1+s)(\kappa_3\kappa_6\kappa_9\kappa_{12})^{\frac{1}{2}} \leq 4(s\kappa_3\kappa_6\kappa_9\kappa_{12})^{\frac{1}{2}}$, then (4.3.8) holds for all $K_1 > 0$. This inequality simplifies to $1 + s \leq 4\sqrt{s}$, which holds if and only if $s \in \big(\frac{14-\sqrt{192}}{2}, \frac{14+\sqrt{192}}{2}\big)$.

Now, inequality (4.3.8) holds for all $K_1 > 0$ if and only if the coefficient of K_1 is nonnegative. We set $r^6 = s$, and the coefficient of K_1 becomes

$$h(r) := -(1+r^6)\kappa_6\kappa_9 \;+\; 3r^2(\kappa_6^2\kappa_9^2 a(\overline{\boldsymbol{\kappa}}))^{\frac{1}{3}}(1+r^2) \;+\; 4r^3(\kappa_3\kappa_6\kappa_9\kappa_{12})^{\frac{1}{2}}.$$

This is a degree 6 polynomial in r with negative leading and independent term and the other coefficients are nonnegative, with at least one positive. Since the right hand side of (4.3.8) evaluated at $s = 1$ is strictly positive, $h(1) > 0$ and h has exactly two distinct positive roots r_1 and r_2. These give rise to two values $s_1(\boldsymbol{\eta'}) = r_1^6, s_2(\boldsymbol{\eta'}) = r_2^6$, satisfying $s_1(\boldsymbol{\eta'}) < \frac{14-\sqrt{192}}{2}$ and $\frac{14+\sqrt{192}}{2} < s_2(\boldsymbol{\eta'})$ for any $\boldsymbol{\eta'}$, and such that (4.3.8) holds for any $s \in [s_1(\boldsymbol{\eta'}), s_2(\boldsymbol{\eta'})]$. This proves (i).

If $s \notin [s_1(\boldsymbol{\eta'}), s_2(\boldsymbol{\eta'})]$, then $h(\sqrt[6]{s})$ is negative, and hence inequality (4.3.8) only holds for $K_1 \leq K_1'$ for $K_1' > 0$ making the right-hand side of (4.3.8) zero. This concludes the proof of (ii).

Finally, (iii) follows from the fact that $a(\overline{\boldsymbol{\kappa}})$ increases with the product $\kappa_3\kappa_{12}$, and hence the positive terms of $h(r)$ also increase. □

4.3.3 Regions of Multistationarity

In this subsection we first point out that multistationarity can indeed be enabled for some $\boldsymbol{\eta} \in \mathbb{R}^8_{>0}$ such that $a(\boldsymbol{\eta}) > 0$ and $b(\boldsymbol{\eta}) < 0$, and describe a specific method to obtain some points in the parameter space that gives rise to multistationarity by fixing $\boldsymbol{\eta'} = (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$. More specifically, we show that if K_4 also is fixed, then multistationarity is enabled for K_1 large enough, and, symmetrically (see Remark 4.2.3), if K_1 is fixed, then K_4 large enough yields multistationarity. This means for a fixed $\boldsymbol{\eta'}$, in the undecided regions near the K_1- and K_4-axes, which were left open by Corollary 4.3.10, multistationarity may occur. Afterwards, we prove Lemma 4.3.13, which essentially points out that each multistationarity region near the axes in the positive orthant of (K_1, K_4)-plane are full dimensional. Later on, this lemma leads to Theorem 4.3.15, which gives an explicit parametric description of the regions of mono- and multistationarity.

Multistationarity can be enabled when $b(\boldsymbol{\eta}) < 0$:

Once $\boldsymbol{\eta'} = (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$, and $K_4 > 0$ are fixed, $p_{\boldsymbol{\eta},H}(x_1, x_3)$ becomes a polynomial in K_1, x_1, x_3 which we denote as $P_{\boldsymbol{\eta'},K_4}(K_1, x_1, x_3)$. Under the hypothesis $a(\overline{\boldsymbol{\kappa}}) \geq 0$ (which is independent of K_1 and K_4), the coefficient of $K_1^2 x_1^2 x_3$ is negative and equals to $-K_2 K_3 \kappa_3 \kappa_6^2 \kappa_9 \kappa_{12}$. The Newton polytope of $P_{\boldsymbol{\eta'},K_4}(K_1, x_1, x_3)$ depends on whether $a(\overline{\boldsymbol{\kappa}}) = 0$ or $a(\overline{\boldsymbol{\kappa}}) > 0$, but in both cases the point $(2, 2, 1)$ is a vertex. Therefore, following the proof of Proposition 4.2.7, we can find specific values for K_1, x_1 and x_3 so that $P_{\boldsymbol{\eta'},K_4}(K_1, x_1, x_3) < 0$. In Proposition 4.3.11, we give a formal description of a nonempty subset that consists of values of K_1 that enables multistationarity for fixed $\boldsymbol{\eta'} = (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ and $K_4 > 0$.

Proposition 4.3.11. Assume $\boldsymbol{\eta'} = (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ and $K_4 > 0$ are fixed and let $\operatorname{int}(\mathscr{N})$ be the interior of the outer normal cone of $\operatorname{New}(P_{\boldsymbol{\eta'},K_4})$ at $(2, 2, 1)$. If K_1 belongs to the set

$$\bigcup_{\boldsymbol{w} \in \mathscr{N}^o} \{y \mid y > z_0^{w_1}, \text{ with } z_0 \text{ the largest root of } P_{\boldsymbol{\eta'},K_4}(z^{w_1}, z^{w_2}, z^{w_3})\}, \tag{4.3.9}$$

then $p_{\boldsymbol{\eta},H}$ attains negative values over $\mathbb{R}^2_{>0}$ and $\boldsymbol{\eta}$ enables multistationarity. Moreover, this set is nonempty. Analogously, by applying the symmetry in Remark 4.2.3 to the polynomial $P_{\boldsymbol{\eta'},K_4}(K_1, x_1, x_3)$, we obtain a set of values of K_4 that enable multistationarity.

Proof. As the point $(2, 2, 1)$ is a vertex of $\operatorname{New}(P_{\boldsymbol{\eta'},K_4})$, there exist $K_1, x_1, x_3 > 0$ such that $P_{\boldsymbol{\eta'},K_4}(K_1, x_1, x_3) < 0$ by Proposition 4.2.7. Following the proof of Proposition 4.2.7, for $\boldsymbol{w} \in \mathscr{N}^o$, we consider the univariate function $P_{\boldsymbol{\eta'},K_4}(z) = P_{\boldsymbol{\eta'},K_4}(z^{w_1}, z^{w_2}, z^{w_3})$, which is a generalized polynomial with real exponents and negative leading term. Then $P_{\boldsymbol{\eta'},K_4}(z) < 0$

for all $z > z_0$, where z_0 is the largest real root of $P_{\boldsymbol{\eta}',K_4}(z)$. This means, the point $\boldsymbol{\eta} = (z^{w_1}, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ enables multistationarity for all $z > z_0$ since $p_{\boldsymbol{\eta},H}(z^{w_2}, z^{w_3}) = P_{\boldsymbol{\eta}',K_4}(z) < 0$ for $z > z_0$. All that remains is to show that w_1 is positive, to rewrite this condition as $K_1 > z_0^{w_1}$ as in the statement.

The outer normal cone $\mathscr{N}$ of $\mathrm{New}\,(P_{\boldsymbol{\eta}'})$ at $(2,2,1)$ is generated by the vectors

$$\begin{aligned} &\boldsymbol{v_1} := (2,1,0), \quad \boldsymbol{v_2} := (1,0,1), \quad \boldsymbol{v_3} := (2,1,2), \qquad \text{if} \quad a(\boldsymbol{\eta}) > 0, \\ &\boldsymbol{v_1} := (2,1,0), \quad \boldsymbol{v_2} := (1,0,1), \quad \boldsymbol{v_3} := (0,0,1), \qquad \text{if} \quad a(\boldsymbol{\eta}) = 0. \end{aligned} \tag{4.3.10}$$

As any vector in $\mathscr{N}$ is of the form $\boldsymbol{w} = \lambda_1\boldsymbol{v_1} + \lambda_2\boldsymbol{v_2} + \lambda_3\boldsymbol{v_3}$ with $\lambda_i > 0$, we have $w_1 > 0$. This concludes the proof.

Further details about the computations can be found in the supplementary file *SupplementaryInfoThesis.mw*. □

Previously in Example 4.2.17, we used the proof of Proposition 4.2.7 to show that a given point $\boldsymbol{\kappa}$ from the parameter space enables multistationarity. Proposition 4.3.11 further describes, using a similar approach, how to find parameter point $\boldsymbol{\eta}$ that enables multistationarity. In Example 4.3.12, we explicitly calculate such a point that enables multistationarity in the parameter region.

Example 4.3.12. Let us fix $\boldsymbol{\eta}' = (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}) = (1,1,2,1,1,1)$ and K_4 run free for now. Consider the vector $\boldsymbol{w} = (3,1,2) = \frac{1}{2}\boldsymbol{v_1} + \boldsymbol{v_2} + \frac{1}{2}\boldsymbol{v_3} \in \mathscr{N}^o$, where $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}$ are given as in (4.3.10). Then,

$$P_{\boldsymbol{\eta}',K_4}(z^3, z, z^2) = z^7(-2z^3 + (7 + 4K_4)z^2 + (14K_4 + 1)z + 14 - 2K_4), \tag{4.3.11}$$

and if we consider the reparameterization $K_1 = z^3$, then (4.3.11) becomes

$$-2K_1^{10/3} + 14K_4K_1^{8/3} + 6K_1^{7/3} + K_1^{8/3} + 7K_1^3 + 4K_4K_1^3 - 2\,(K_4 - 4)\,K_1^{7/3}. \tag{4.3.12}$$

For example, if we fix $K_4 = 1$, then the largest root of $P_{\boldsymbol{\eta}',K_4}(z)$ in (4.3.11) is $z_0 \approx 6.75$. Therefore, if $z > z_0$, that is, if $K_1 = z^3 > z_0^3$, then multistationarity is enabled since $p_{\boldsymbol{\eta},H}(z, z^2) < 0$ for $\boldsymbol{\eta} = (z^3, 1, 1, 1, 2, 1, 1, 1)$. We can easily verify this: e.g., if we set $z = 7$, then $p_{\boldsymbol{\eta},H}(7, 49) = -24706290 < 0$ for $\boldsymbol{\eta} = (343, 1, 1, 1, 2, 1, 1, 1)$.

Similarly, any given $K_4 > 0$ yields a new univariate $P_{\boldsymbol{\eta}',K_4}(z)$ with a new largest root. Then, we can choose a value for $K_1 > z_0^3$ for each $K_4 > 0$ to obtain the solid red curve in Figure 4.9. The dashed green curves are computed via the same method, but we use $\boldsymbol{w} = (\frac{1}{2}, 2, 1)$ and $\boldsymbol{w} = (\frac{1}{3}, 3, 2)$ as vectors in $\mathscr{N}$. ⬡

Obtaining an explicit description of the monostationarity region in Proposition 4.3.11, in terms of algebraic inequalities in the parameters has not been possible with using only

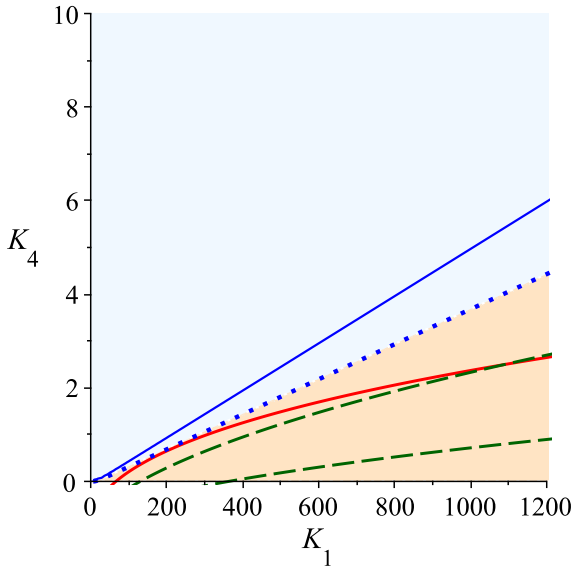

Figure 4.9: With $(K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}) = (1, 1, 2, 1, 1, 1)$, the figure shows a dotted blue line separating the regions of monostationarity (above the line, blue) and of multistationarity (below the line, orange), found from Theorem 4.3.15. Above the solid blue line in the monostationarity region, the condition in Theorem 4.3.5 is satisfied. Below the solid red line in the multistationarity region, multistationarity is enabled by means of Proposition 4.3.11 with $\boldsymbol{w} = (3, 1, 2) \in \mathscr{N}$; similarly, the green dashed lines correspond to $\boldsymbol{w} = (\frac{1}{2}, 2, 1)$ and $\boldsymbol{w} = (\frac{1}{3}, 3, 2)$.

Proposition 4.3.11. However, we address this issue in the next subsection after we prove some auxiliary facts. In particular, we provide an explicit parametric description of the region of multistationarity, which gives rise to the dotted blue line in Figure 4.9.

Parameterization of the region of multistationarity:

As before, we let $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ denote a point in the parameter space $\mathbb{R}^8_{>0}$ such that $a(\overline{\kappa}) \geq 0$, and we denote the vector obtained by forgetting first and fourth entry of $\boldsymbol{\eta}$ as $\boldsymbol{\eta'} = (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$. In this subsection, we describe three functions

$$\begin{array}{lll} \psi : \mathbb{R}^7_{>0} \to \mathbb{R}_{>0}, & \phi : \mathbb{R}^8_{>0} \to \mathbb{R}_{>0}, & \xi : \mathbb{R}^6_{>0} \to \mathbb{R}_{>0} \\ (s, \boldsymbol{\eta'}) \mapsto \psi(s, \boldsymbol{\eta'}) & (s, \psi(s, \boldsymbol{\eta'}), \boldsymbol{\eta'}) \mapsto \phi(s, \psi(s, \boldsymbol{\eta'}), \boldsymbol{\eta'}) & \boldsymbol{\eta'} \mapsto \xi(\boldsymbol{\eta'}) \end{array} \tag{4.3.13}$$

such that $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ enables multistationarity if and only if

$$K_1 = \psi(s, \boldsymbol{\eta'}), \quad K_4 > \phi(s,\, \psi(s, \boldsymbol{\eta'})\,, \boldsymbol{\eta'}), \quad \text{for } s \in (0, \xi(\boldsymbol{\eta'})), \quad \text{or} \quad (4.3.14)$$

$$K_4 = \psi(s, \sigma(\boldsymbol{\eta'})), \quad K_1 > \phi(s,\, \psi(s, \sigma(\boldsymbol{\eta'})\,), \sigma(\boldsymbol{\eta'})), \quad \text{for } s \in (0, \xi(\sigma(\boldsymbol{\eta'})). \quad (4.3.15)$$

Note that if $\boldsymbol{\eta'}$ is fixed, then Proposition 4.3.11 and Corollary 4.3.10, together with the fact that $b(\boldsymbol{\eta}) > 0$ for K_1, K_4 small, indicate that there are two branches of multistationarity along the two axes: one with K_1 large and K_4 small, and one with K_4 large and K_1 small. These are the two branches giving rise to the two conditions (4.3.14) and (4.3.15). By the symmetry of the system, we describe the K_4-branch given by (4.3.14), and the other branch results from applying σ. We specify the nature of these branches further in the following lemma.

Lemma 4.3.13. Assume that $\boldsymbol{\eta^*} = (K_1^*, K_2, K_3, K_4^*, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ enables multistationarity and $a(\overline{\boldsymbol{\kappa}}) \geq 0$. Then, the following statements hold:

(i) If $K_1^* < K_4^*$, then for all $K_4 \geq K_4^*$ and $K_1 \leq K_1^*$ the parameter point $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ also enables multistationarity.

(ii) If $K_1^* > K_4^*$, then for all $K_4 \leq K_4^*$ and $K_1 \geq K_1^*$ the parameter point $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ also enables multistationarity.

Remark 4.3.14. Due to Corollary 4.3.10, any vector $\widetilde{\boldsymbol{\eta}} = (K_1^*, K_2, K_3, K_1^*, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ precludes multistationarity, and thus the case $K_1^* = K_4^*$ is omitted in Lemma 4.3.13. This fact will also play a crucial role in the proof of Lemma 4.3.13. ⬡

Proof of Lemma 4.3.13. As $\boldsymbol{\eta^*}$ enables multistationarity, there exist $x_1, x_3 > 0$ such that $p_{\boldsymbol{\eta^*},H}(x_1, x_3) < 0$. We fix these values of x_1, x_3 along with the parameters $K_1 = K_1^*$, and $\boldsymbol{\eta'} = (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ of $\boldsymbol{\eta^*}$. The crucial observation is that $p_{\boldsymbol{\eta^*},H}$, with $\boldsymbol{\eta'}, x_1, x_3, K_1 = K_1^*$ fixed, is simply a linear polynomial $q(K_4) = c_1 K_4 + c_0$ in K_4.

On the one hand, $q(K_4^*) < 0$, because $q(K_4^*) = p_{\boldsymbol{\eta^*},H}(x_1, x_3) < 0$. On the other hand, $q(K_1^*) \geq 0$, since $q(K_1^*) = p_{\widetilde{\boldsymbol{\eta}},H}(x_1, x_3)$ where $\widetilde{\boldsymbol{\eta}} = (K_1^*, K_2, K_3, K_1^*, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$, and $\widetilde{\boldsymbol{\eta}}$ precludes multistationarity due to Corollary 4.3.10. This means that $c_1 \neq 0$. First, if $K_4^* > K_1^*$ holds, then $q(K_4^*) - q(K_1^*) = c_1(K_4^* - K_1^*) < 0$. This implies that $c_1 < 0$, and consequently $q(K_4) < 0$ must hold for any $K_4 \geq K_4^*$. If $K_4^* < K_1^*$ holds, then it similarly implies that $c_1 > 0$. Thus, c_0 is necessarily negative, which means that $q(K_4) < 0$ for any $K_4 \leq K_4^*$.

Therefore, if $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ enables multistationarity, then the inequalities in the statement regarding K_4 hold. The inequalities for K_1 follow by using the symmetry in Remark 4.2.3. □

Based on Lemma 4.3.13, we define the *K_4-branch of multistationarity* to consist of the set of parameters $\boldsymbol{\eta} = (K_1^*, K_2, K_3, K_4^*, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ enabling multistationarity and such that $K_4^* > K_1^*$. If $\boldsymbol{\eta}$ is a point in this branch, then $\widetilde{\boldsymbol{\eta}} = (K_1^*, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ also enables multistationarity for all $K_4 \geq K_4^*$. For fixed parameters $K_1^*, K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}$, we wish to determine the *infimum* value K_4^* that satisfies this property, that is, the value K_4^* such that for any $K_4 > K_4^*$ multistationarity is enabled.

In the next theorem we identify this value parameterically: we give functions $\psi(s, \boldsymbol{\eta'})$ and $\phi(s, K_1, \boldsymbol{\eta'})$, for s in an interval of the form $(0, \xi(\boldsymbol{\eta'}))$, such that for any $K_4 > \phi(s, \psi(s, \boldsymbol{\eta'}), \boldsymbol{\eta'})$, the point $(\psi(s, \boldsymbol{\eta'}), K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ enables multistationarity, but for $K_4 \leq \phi(s, \psi(s, \boldsymbol{\eta'}), \boldsymbol{\eta'})$, multistationarity is not enabled. For fixed $\boldsymbol{\eta'}$, the pair $(\psi(s, \boldsymbol{\eta'}), \phi(s, \psi(s, \boldsymbol{\eta'}), \boldsymbol{\eta'}))$ describes a curve in the (K_1, K_4)-plane separating the region of monostationarity and multistationarity along the K_4-branch. The K_1-branch of multistationarity is defined analogously.

Before we give the rigorous definitions of ψ, ϕ and ξ, we first define the following auxiliary functions in s, K_1 and $\boldsymbol{\eta'}$:

$$\begin{aligned}
\alpha_1(s, \boldsymbol{\eta'}) &= -K_2\kappa_3^2 s^3\Big(K_2(K_2+K_3)\kappa_3\kappa_9\kappa_{12}s + K_3\kappa_{12}\big(2K_2 a(\overline{\boldsymbol{\kappa}}) + (K_2+K_3)\kappa_3\kappa_{12}\big) \\
&\quad + \sqrt{K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}})}\,\big(2K_2\kappa_9 s + K_2\kappa_{12} + 3\,K_3\kappa_{12}\big)\Big), \\
\beta_1(s, \boldsymbol{\eta'}) &= \kappa_6\Big(- K_2^2\kappa_3^2\kappa_9^2 s^4 + K_2\kappa_3^2\kappa_9\kappa_{12}s^3(3K_2 - K_3) + 2K_2K_3\kappa_3\kappa_{12}s^2(4\kappa_3\kappa_{12} - \kappa_9\kappa_6) \\
&\quad - K_3\kappa_3\kappa_6\kappa_{12}^2 s(K_2 - 3K_3) - K_3^2\kappa_6^2\kappa_{12}^2 \\
&\quad + 2s\sqrt{K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}})}\big(K_2\kappa_3\kappa_9 s^2 + 2(K_2+K_3)\kappa_3\kappa_{12}s + K_3\kappa_6\kappa_{12}\big)\Big),
\end{aligned}$$

and

$$\begin{aligned}
\alpha_4(s, K_1, \boldsymbol{\eta'}) &= K_3\kappa_{12}\Big(K_2\kappa_3 s^2\big(K_1\kappa_6\kappa_9 - (K_2+K_3)\kappa_3\kappa_{12}\big) - 2K_1K_2\kappa_3\kappa_6\kappa_{12}s - K_1K_3\kappa_6^2\kappa_{12} \\
&\quad - 2s\sqrt{K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}})}\,\big(K_2\kappa_3 s + K_1\kappa_6\big)\Big), \\
\beta_4(s, \boldsymbol{\eta'}) &= K_2\kappa_3\kappa_9 s^2\Big(2s\sqrt{K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}})} + K_2\kappa_3\kappa_9\,s^2 + 2K_3\kappa_3\kappa_{12}\,s - K_3\kappa_6\kappa_{12}\Big).
\end{aligned}$$

We let now

$$\psi(s, \boldsymbol{\eta'}) = \frac{\alpha_1(s, \boldsymbol{\eta'})}{\beta_1(s, \boldsymbol{\eta'})}, \qquad \phi(s, K_1, \boldsymbol{\eta'}) = \frac{\alpha_4(s, K_1, \boldsymbol{\eta'})}{\beta_4(s, \boldsymbol{\eta'})}, \tag{4.3.16}$$

and let $\xi(\boldsymbol{\eta'})$ be the first positive root of the polynomial $\beta_1(s, \boldsymbol{\eta'}) \in \mathbb{R}[s]$ for the fixed $\boldsymbol{\eta'}$.

We also note that the proof of Theorem 4.3.15 uses the function `IsEmpty` in `Maple 2019`, which checks whether the zero set of a finite collection of polynomials is empty or not. The details of the computation can be found on the supplementary file *Supplemen-*

taryInfoThesis.mw in the end of the thesis, or available in the following link:

https://moto.math.nat.tu-bs.de/appliedalgebra_public/oguzhan_yuruk_thesis_supplementary_file

Theorem 4.3.15. Let $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}) \in \mathbb{R}^8_{>0}$ such that $a(\boldsymbol{\eta}) \geq 0$, the map σ be defined as in Remark 4.2.3, and as before $\boldsymbol{\eta'} = (K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ denote the fixed entries of $\boldsymbol{\eta}$. Then, multistationarity is enabled if and only if K_1, K_4 are as in one of the following cases:

$$K_1 = \psi(s, \boldsymbol{\eta'}), \quad \text{and} \quad K_4 > \phi(s, \psi(s, \boldsymbol{\eta'}), \boldsymbol{\eta'}), \qquad \text{with} \quad s \in (0, \xi(\boldsymbol{\eta'})),$$

or

$$K_4 = \psi(s, \sigma(\boldsymbol{\eta'})), \quad \text{and} \quad K_1 > \phi(s, \psi(s, \sigma(\boldsymbol{\eta'})), \sigma(\boldsymbol{\eta'})), \qquad \text{with} \quad s \in (0, \xi(\sigma(\boldsymbol{\eta'})).$$

The first case describes the K_4-branch, and $s = x_1$, while the second case describes the K_1-branch, and $s = x_2$. Furthermore, for any $\boldsymbol{\eta'}$, ψ is increasing for s in the considered interval and the image this interval under ψ is in $\mathbb{R}_{>0}$.

Proof. We consider $\boldsymbol{\eta'}$ fixed and study the K_4-branch. The proof relies on several symbolic computations that can be found in the accompanying supplementary file *SupplementaryInfoThesis.mw*. Recall from the proof of Lemma 4.3.13 that $p_{\boldsymbol{\eta},H}(x_1, x_3)$ is linear in K_4. If written as $c_1 K_4 + c_0$ we have

$$\begin{aligned}
c_1 &= K_2\kappa_3\kappa_9 x_1^2\Big(K_2\kappa_3 a(\overline{\boldsymbol{\kappa}})x_1^2x_3^2 + K_1\kappa_6\big(K_2\kappa_3\kappa_9x_1^2 + 2\,K_3\kappa_3\kappa_{12}x_1 - K_3\kappa_6\kappa_{12}\big)x_3 + K_1^2K_3\kappa_6^2\kappa_{12}\Big)\\
c_0 &= K_3\kappa_{12}\Big(K_2\kappa_3 a(\overline{\boldsymbol{\kappa}})(K_2\kappa_3x_1 + K_1\kappa_6)x_1^2x_3^2 - K_1\kappa_6\big(K_2\kappa_3(K_1\kappa_6\kappa_9 - (K_2+K_3)\kappa_3\kappa_{12})x_1^2\\
&\quad - K_1\kappa_6\kappa_{12}(2K_2\kappa_3x_1 + K_3\kappa_6)\big)x_3 + K_1^2K_3\kappa_6^2\kappa_{12}(K_2\kappa_3x_1 + K_1\kappa_6)\Big).
\end{aligned}$$

In order to understand the K_4-branch, we consider the case $c_1 < 0$ (see the proof of Lemma 4.3.13). For fixed x_1, x_3, K_1, this implies that the coefficient of x_3 in c_1 is negative, which in turn implies that x_1 is smaller than the positive root of $K_2\kappa_3\kappa_9x_1^2 + 2K_3\kappa_3\kappa_{12}x_1 - K_3\kappa_6\kappa_{12}$, namely, smaller than

$$x_{1,\text{bound}} := \frac{-K_3\kappa_3\kappa_{12} + \sqrt{K_3\kappa_3\kappa_{12}(K_2\kappa_6\kappa_9 + K_3\kappa_3\kappa_{12})}}{K_2\kappa_3\kappa_9}.$$

Under the assumption $x_1 < x_{1,\text{bound}}$, and $a(\overline{\boldsymbol{\kappa}}) \geq 0$, using the function `IsEmpty` in `Maple 2019`, we find that $c_0 > 0$. Hence for $\boldsymbol{\eta}$ in the K_4-branch, if $p_{\boldsymbol{\eta},H}(x_1, x_3) < 0$, then necessarily $c_1 < 0$ and $c_0 > 0$. Furthermore, for $\boldsymbol{\eta}$ in the K_4-branch, $p_{\boldsymbol{\eta},H}(x_1, x_3) = 0$ holds if and only if $K_4 = \frac{-c_0}{c_1} > 0$, and $p_{\boldsymbol{\eta},H}(x_1, x_3) < 0$ holds if $K_4 > \frac{-c_0}{c_1}$. It follows that

the boundary of the K_4-branch is determined by minimizing $\frac{-c_0}{c_1}$ with respect to $x_1, x_3 > 0$ subject to $c_1 < 0$. For $a(\overline{\boldsymbol{\kappa}}) > 0$, we find the minimum value of $\frac{-c_0}{c_1}$, and for $a(\overline{\boldsymbol{\kappa}}) = 0$, we find its infimum value.

For a fixed $x_1 > 0$, we consider first $\frac{-c_0}{c_1}$ as a function of x_3 in the region where $c_1 < 0$. When $a(\overline{\boldsymbol{\kappa}}) > 0$, the derivative has a unique positive zero at

$$x_{3,\min} := \frac{K_1\kappa_6\sqrt{K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}})}}{K_2\kappa_3 a(\overline{\boldsymbol{\kappa}})x_1},$$

which defines a minimum. We evaluate $\frac{-c_0}{c_1}$ at $x_{3,\min}$, which now becomes the function $\phi(x_1, K_1, \boldsymbol{\eta'})$ in (4.3.16). When $a(\overline{\kappa}) = 0$, $\frac{-c_0}{c_1}$ is strictly decreasing, and hence the infimum value it attains is the limit as x_3 goes to $+\infty$, which is $\phi(x_1, K_1, \boldsymbol{\eta'})$ again. It makes sense then to set $x_{3,\min} = +\infty$ in this case. Hence $\phi(x_1, K_1, \boldsymbol{\eta'})$ gives, for fixed $\boldsymbol{\eta'}$, K_1, and x_1 such that $c_1 < 0$, the minimal/infimum value of $\frac{-c_0}{c_1}$ seen as a function of x_3.

We notice that the denominator of $\phi(x_1, K_1, \boldsymbol{\eta'})$ (which is a multiple of $c_1(x_1, x_{3,\min})$ when $a(\overline{\boldsymbol{\kappa}}) > 0$), is a polynomial in x_1 of the form x_1^2 times a quadratic polynomial. The latter has positive leading term and negative independent term. Hence it has a unique positive root γ (which we can compute), and this denominator is negative if and only if $x_1 \in (0, \gamma)$. When $a(\overline{\boldsymbol{\kappa}}) = 0$, we have $\gamma = x_{1,\text{bound}}$.

In particular ϕ is continuous and differentiable in $(0, \gamma)$. The function ϕ is a rational function in x_1 of the following form:

$$\phi(x_1, K_1, \boldsymbol{\eta'}) = \frac{a_1x_1^2 - a_2x_1 - a_3}{x_1^2(b_1x_1^2 + b_2x_1 - b_3)},$$

where a_2, a_3, b_1, b_2, b_3 depend on $\boldsymbol{\eta'}, K_1$ and are positive under the current hypotheses, and a_1, which also depends on $K_1, \boldsymbol{\eta'}$ is

$$a_1 := -K_2K_3\kappa_3\kappa_{12}\Big(\big(K_1\kappa_6\kappa_9 - (K_2 + K_3)\kappa_3\kappa_{12}\big) - 2\sqrt{K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}})}\Big).$$

In order to minimize ϕ in $(0, \gamma)$, we find the derivative of ϕ with respect to x_1:

$$\phi'(x_1, K_1, \boldsymbol{\eta'}) := \frac{d\phi}{dx_1}(x_1, K_1, \boldsymbol{\eta'}) = \frac{-2a_1b_1x_1^4 + (3a_2b_1 - a_1b_2)x_1^3 + (2a_2b_2 + 4a_3b_1)x_1^2 + (3a_3b_2 - a_2b_3)x_1 - 2a_3b_3}{x_1^3(b_1x_1^2 + b_2x_1 - b_3)^2}.$$

The extreme values of ϕ' are determined by the zeroes of its numerator. This numerator is a polynomial $u(x_1)$ in x_1 with negative constant term and positive degree 2 term. If $a_1 \leq 0$, then the leading and degree 3 coefficients of $u(x_1)$ are nonnegative. By Descartes' rule of signs, it follows that $\phi' = 0$ has exactly one positive root, which, in case that it belongs to $(0, \gamma)$, gives rise to a minimum of ϕ, as the independent term of the numerator of ϕ' is negative.

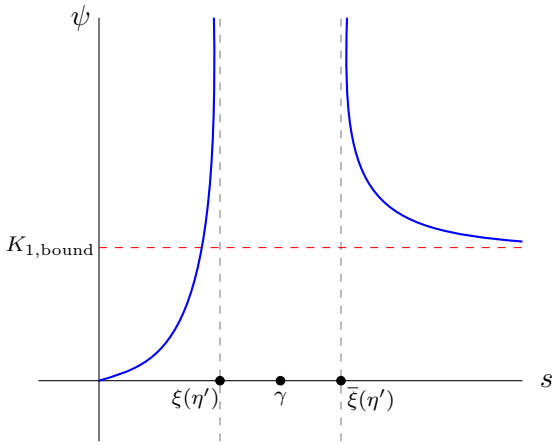

Figure 4.10: Cartoon depiction of the function $\psi(s, \boldsymbol{\eta'})$ for a fixed $\boldsymbol{\eta'}$, with $\xi(\boldsymbol{\eta'}), \overline{\xi}(\boldsymbol{\eta'})$, γ and $K_{1,\text{bound}}$ as given in the proof of Theorem 4.3.15.

If $a_1 > 0$, then the leading term of $u(x_1)$ is negative, and by the Descartes' rule of signs, $\phi' = 0$ at most two positive roots, in which case the first positive root is a minimum of ϕ if it belongs to $(0, \gamma)$ as above. Note that $a_1 > 0$ if and only if

$$K_1 > K_{1,\text{bound}}, \qquad \text{where} \quad K_{1,\text{bound}} := \frac{(K_2+K_3)\kappa_3\kappa_{12}+2\sqrt{K_2K_3\kappa_3\kappa_{12}a(\overline{\boldsymbol{\kappa}})}}{\kappa_6\kappa_9}.$$

The next step is thus to confirm that the only positive root in the case $a_1 \leq 0$ is smaller than γ, and that there is such a (simple) positive root in the case $a_1 > 0$. To this end, we observe that the numerator of ϕ' is linear in K_1. By solving the numerator for K_1, we obtain that any extreme value satisfies $K_1 = \psi(x_1, \boldsymbol{\eta'})$ with ψ as in (4.3.16). The denominator $\beta_1(x_1, \boldsymbol{\eta'})$ has degree 4 in x_1, negative leading and constant terms, and the coefficient of x_1^2 is positive. By Descartes' rule of signs, $\beta_1(x_1, \boldsymbol{\eta'})$ has at most two positive roots. Using the function `IsEmpty` in `Maple 2019`, we find that $\beta_1(\gamma, \boldsymbol{\eta'}) > 0$. This implies that $\beta_1(x_1, \boldsymbol{\eta'})$ has exactly one simple positive root $\xi(\boldsymbol{\eta'})$ in the interval $(0, \gamma)$ and one simple positive root $\overline{\xi}(\boldsymbol{\eta'})$ in $(\gamma, +\infty)$. The numerator $\alpha_1(x_1, \boldsymbol{\eta'})$ of ψ has degree 4 in x_1, is negative for $x_1 > 0$, and vanishes at $x_1 = 0$. Hence, $\psi(x_1, \boldsymbol{\eta'})$ is positive in the intervals $(0, \xi(\boldsymbol{\eta'}))$ and $(\overline{\xi}(\boldsymbol{\eta'}), +\infty)$. It tends to infinity when x_1 tends to $\xi(\boldsymbol{\eta'})$ from the left and also to $\overline{\xi}(\boldsymbol{\eta'})$ from the right. Furthermore, ψ vanishes at $x_1 = 0$ and tends to $K_{1,\text{bound}}$ when x_1 tends to infinity. In particular, the image of ψ over the interval $(0, \xi(\boldsymbol{\eta'}))$ is $\mathbb{R}_{>0}$, and the image over the interval $(\overline{\xi}(\boldsymbol{\eta'}), +\infty)$ is $(K_{1,\text{bound}}, +\infty)$. See Figure 4.10. The image of $(\xi(\boldsymbol{\eta'}), \overline{\xi}(\boldsymbol{\eta'}))$ by ψ belongs to $\mathbb{R}_{<0}$.

The preimages of a given K_1 by ψ are the zeroes of $\phi' = 0$. By comparing the image of ψ to the discussion on the sign of a_1 and the positive roots of ϕ' above, we conclude that ψ is strictly increasing in $(0, \xi(\boldsymbol{\eta'}))$, and each x_1 in this interval such that $K_1 = \psi(x_1, \boldsymbol{\eta'})$

is a simple root of $\phi' = 0$. In particular, ϕ attains its minimum at the preimage of K_1 by ψ in the interval $(0, \gamma)$.

To summarize, we have shown that given $K_1 > 0$, and $\overline{x}_1 \in (0, \xi(\boldsymbol{\eta'}))$ such that $K_1 = \psi(\overline{x}_1, \boldsymbol{\eta'})$, K_4 gives rise to a parameter point enabling multistationarity in the K_4-branch if and only if K_4 is larger than $\frac{-c_0}{c_1}$ evaluated at $x_{3,\min}$ and $\overline{x}_1$, where we already know that $c_1 < 0$ as $\xi(\boldsymbol{\eta'}) < \gamma$. This gives that $\boldsymbol{\eta}$ enables multistationarity in the K_4-branch if and only if there exists $x_1 \in (0, \xi(\boldsymbol{\eta'}))$ such that $K_1 = \psi(x_1, \boldsymbol{\eta'})$ and $K_4 > \phi(x_1, \psi(x_1, \boldsymbol{\eta'}), \boldsymbol{\eta'})$. This concludes the proof for K_4-branch; the proof of K_1-branch follows by symmetry using Remark 4.2.3. □

For $\boldsymbol{\eta'}$ fixed, as in Figure 4.9, we obtain a polynomial in K_1, K_4 whose zero set includes the dotted blue curve in Figure 4.9 given by the parameterization, as well as additional components. The blue dotted curve in Figure 4.9 shows the K_1-branch of the multistationarity region given in Theorem 4.3.15 when $(K_2, K_3, \kappa_3, \kappa_6, \kappa_9, \kappa_{12}) = (1, 1, 2, 1, 1, 1)$.

Example 4.3.16. Theorem 4.3.15 provides a method to verify whether a given $\boldsymbol{\eta}$ enables multistationarity. First, we decide if one can verify the monostationarity by utilizing Theorem 4.3.5. If not, and $K_4 > K_1$, then determine $s \in (0, \xi(\boldsymbol{\eta'}))$ such that $K_1 = \psi(s, \boldsymbol{\eta'})$ for $s \in (0, \xi(\boldsymbol{\eta'}))$, and decide whether $K_4 > \phi(s, \psi(s, \boldsymbol{\eta'}), \boldsymbol{\eta'})$. If $K_1 > K_4$, use the expressions for the K_1-branch.

For example, let $\boldsymbol{\eta} = (3, 1, 1, 700, 2, 1, 1, 1)$. Inequality (4.3.4) in Theorem 4.3.5 does not hold. As $K_4 > K_1$, we consider the K_4-branch. We solve $3 = \psi(s, \boldsymbol{\eta'})$ for $s \in (0, \xi(\boldsymbol{\eta'}))$ and obtain $s \approx 0.174$, which gives $\phi(s, \psi(s, \boldsymbol{\eta'}), \boldsymbol{\eta'}) \approx 818.17$. As $700 < 818.17$, the given parameter point does not enable multistationarity. Via a similar computation, it follows as well that the parameter point $(3, 1, 1, 900, 2, 1, 1, 1)$ enables multistationarity. ⬡

4.3.4 Connectivity

In this section we show that the open set $X \subseteq \mathbb{R}^8_{>0}$ of parameter points that enable multistationarity is connected. As any $\boldsymbol{\eta} \in \mathbb{R}^8_{>0}$ either enables or precludes multistationarity, the set $\mathbb{R}^8_{>0} \setminus X$ consists of the parameter points that preclude multistationarity.

We consider X as a topological subspace of $\mathbb{R}^8_{>0}$ with the Euclidean topology. We start by highlighting in the next lemma a path connected subset of X. Recall from (4.2.13) that $a(\boldsymbol{\eta}) = \kappa_3\kappa_{12} - \kappa_6\kappa_9$, and let $Y \subseteq \mathbb{R}^8_{>0}$ consist of the parameter points $\boldsymbol{\eta}$ such that $a(\boldsymbol{\eta}) < 0$.

Lemma 4.3.17. The following subsets of $\mathbb{R}^4$ are path connected:

$$A_{<0} = \{\overline{\boldsymbol{\kappa}} = (\kappa_3, \kappa_6, \kappa_9, \kappa_{12}) \in \mathbb{R}^4_{>0} \mid a(\overline{\kappa}) < 0\}, \quad A_{\geq 0} = \{\overline{\boldsymbol{\kappa}} = (\kappa_3, \kappa_6, \kappa_9, \kappa_{12}) \in \mathbb{R}^4_{>0} \mid a(\overline{\kappa}) \geq 0\}.$$

Additionally, Y is path connected.

Proof. Consider the continuous map $h\colon \mathbb{R}^4_{>0} \to \mathbb{R}^2_{>0}$ sending $(\kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ to $(\kappa_3\kappa_{12}, \kappa_6\kappa_9)$. The fibers of this map are path connected. As $A_{<0}$ and $A_{\geq 0}$ are respectively the preimages by h of the path connected subsets $\{\boldsymbol{x} \in \mathbb{R}^2_{>0} \mid x_1 < x_2\}$ and $\{\boldsymbol{x} \in \mathbb{R}^2_{>0} \mid x_1 \geq x_2\}$ of $\mathbb{R}^2_{>0}$, they are also path connected. Y is also path connected as it is homeomorphic to $\mathbb{R}^4_{>0} \times A_{<0}$. □

By Proposition 4.2.15, multistationarity is enabled whenever $a(\boldsymbol{\eta}) < 0$. Therefore, Y is a subset of X. To show that X is path connected it is enough to show that there exists a path from any point in X to a point in Y.

Theorem 4.3.18. *X and $\mathbb{R}^8_{>0} \setminus X$ are path connected.*

Proof. We start by showing that X is path connected. Let $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa'_{12}) \in X$ such that $a(\boldsymbol{\eta}) \geq 0$. By Lemma 4.3.17, it is enough to show that there exists a path in X that connects $\boldsymbol{\eta}$ to a point $\boldsymbol{\eta}^* \in Y$. As $\boldsymbol{\eta} \in X$ enables multistationarity and $a(\boldsymbol{\eta}) \geq 0$, we can choose $z_1, z_3 > 0$ such that $p_{\boldsymbol{\eta},H}(z_1, z_3) < 0$ due to (ii) in Proposition 4.2.15. We let $\boldsymbol{\eta'} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9)$ and let $P_{\boldsymbol{\eta'},H}(x_1, x_3, \kappa_{12})$ denote $p_{\boldsymbol{\eta},H}$ seen as a polynomial in x_1, x_3, κ_{12}. The vertices of the Newton polytope of $P_{\boldsymbol{\eta'},H}$ are (see Figure 4.11):

$$\operatorname{Vert}\left(\operatorname{New}\left(P_{\boldsymbol{\eta'},H}\right)\right) = \{(0,1,2),(2,2,1),(2,2,2),(1,0,2),(2,0,1),(0,0,2),(3,2,2),(4,2,1),(4,1,0),(4,2,0)\}.$$

The coefficients of the vertices $(2,2,1)$ and $(4,2,0)$ are negative. These two vertices lie on the one dimensional face F given by the intersection of the supporting hyperplanes $x_3 - 2 = 0$ and $-x_1 - 2\kappa_{12} + 4 = 0$. Therefore, the outer normal cone at F is generated by the vectors $v_1 := (0,1,0)$ and $v_2 := (-1,0,-2)$. Following the proof of Proposition 4.2.7, we consider $w := v_1 + v_2 = (-1,1,-2)$ and evaluate $P_{\boldsymbol{\eta'},H}$ at $(z_1 s^{-1}, z_3 s, \kappa'_{12} s^{-2})$. The denominator is positive and the numerator is

$$\begin{aligned} q(s) := &- K_2\kappa_3\kappa_6\kappa_9 z_1^2 z_3^2 (K_2 K_4 \kappa_3\kappa_9 z_1^2 + K_1 K_3 \kappa_6 \kappa'_{12}) s^3 + \kappa_6 z_3 (K_2^2 \kappa_3^2 \kappa_9 (K_1 K_4 \kappa_9 z_1^4 - K_3 \kappa'_{12} z_1^3 z_3) \\ &- K_1 K_2 K_3 \kappa_3 \kappa_6 \kappa_9 \kappa'_{12} (K_1 + K_4) z_1^2 + K_1^2 K_3^2 \kappa_6^2 \kappa_{12}'^2) s^2 + \big(K_2 K_4 \kappa_3 \kappa_9 \kappa'_{12} z_1^2 (K_2 \kappa_3^2 z_1^2 z_3^2 \\ &+ 2K_1 K_3 \kappa_3 \kappa_6 z_1 z_3 + K_1^2 K_3 \kappa_6^2) + K_1 K_3 \kappa_6 \kappa_{12}'^2 (K_2 \kappa_3^2 z_1^2 z_3^2 + 2K_1 K_2 \kappa_3 \kappa_6 z_1 z_3 + K_1^2 K_3 \kappa_6^2)\big) s \\ &+ K_2 K_3 \kappa_3 \kappa_{12}'^2 z_1 \Big(K_2 \kappa_3^2 z_1^2 z_3^2 + K_1 \kappa_3 \kappa_6 (K_2 + K_3) z_1 z_3 + K_1^2 K_3 \kappa_6^2\Big). \end{aligned}$$

We note that for $s = 1$, $q(1) = P_{\boldsymbol{\eta'},H}(z_1, z_3, \kappa'_{12})$ is negative. Further note that the polynomial q has degree 3 in s, its leading coefficient is negative and the coefficients of degree 0 and 1 are positive. By Descartes' rule of signs, q has exactly one positive root, and together with the fact that $q(1) < 0$, it implies $q(s) < 0$ for all $s \geq 1$. Hence, $\boldsymbol{\eta}(s) = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa'_{12} s^{-2}) \in X$ for all $s \geq 1$. As s increases, $\kappa'_{12} s^{-2}$

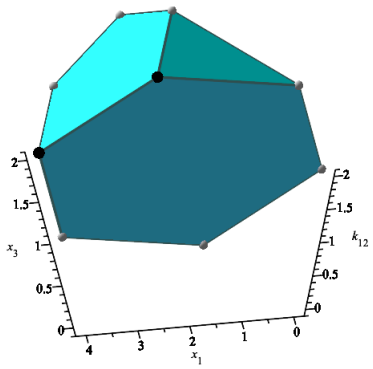

Figure 4.11: Newton Polytope of $P_{\boldsymbol{\eta}',H}$ as a polynomial in x_1, x_3, κ_{12}. In black we show two negative vertices.

decreases and hence $a(\boldsymbol{\eta}(s))$ decreases. For $s > \sqrt{\frac{\kappa_3\kappa'_{12}}{\kappa_6\kappa_9}}$, we have $a(\boldsymbol{\eta}(s)) < 0$ and hence $a(\boldsymbol{\eta}(s)) \in Y$. This provides the desired path, which proves the first part of the statement.

To study $Z := \mathbb{R}^8_{>0} \setminus X$, note that the set of points $\boldsymbol{\eta}$ with $K_1 = K_4$ and $a(\overline{\boldsymbol{\kappa}}) \geq 0$ is path connected by Lemma 4.3.17, and is further a subset of Z by Corollary 4.3.10. By Lemma 4.3.13, in Z there are paths joining any $\boldsymbol{\eta} = (K_1, K_2, K_3, K_4, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$ in Z to $\boldsymbol{\eta}^* = (K_1, K_2, K_3, K_1, \kappa_3, \kappa_6, \kappa_9, \kappa_{12})$. Hence $Z := \mathbb{R}^8_{>0} \setminus X$ is path connected. This concludes the proof of the theorem. □

Remark 4.3.19. According to Theorem 4.3.18, the region X of parameters $\boldsymbol{\eta}$ that enable multistationarity is connected in $\mathbb{R}^8_{>0}$. For this system, the preimage of X by π, that is, the set of parameters $\boldsymbol{\kappa} \in \mathbb{R}^{12}_{>0}$ that enable multistationarity, is also path connected in $\mathbb{R}^{12}_{>0}$. To see this, it is enough to study the map $(\kappa_1, \kappa_2, \kappa_3) \mapsto (\frac{\kappa_2+\kappa_3}{\kappa_1}, \kappa_3)$. The fiber of this map of each point in the image is one dimensional and connected. The map π comprises four disjoint copies of such a map, and hence the fiber by π of a point in the image is four dimensional and connected. Therefore, the preimage of X by π is path-connected. ⬡

4.4 Higher Number of Sites

We recall that [FKdWY20] addresses exclusively to the case of 2-site. However, some aspects of [FKdWY20] are applicable to the phosphorylation cycles with more than 2 sites. In this section we discuss, and point out some preliminary facts about the regions of mono and multistationarity in the cases of 3- and 4-site phosphorylation cycles. The content of this section have not been published in an article yet, since it is still a part of an ongoing research.

In the case of 3-site phosphorylation, the network given in (4.2.1) becomes

$$\begin{aligned} &S_0 + E \underset{\kappa_2}{\overset{\kappa_1}{\rightleftarrows}} ES_0 \xrightarrow{\kappa_3} S_1 + E \underset{\kappa_8}{\overset{\kappa_7}{\rightleftarrows}} ES_1 \xrightarrow{\kappa_9} S_2 + E \underset{\kappa_{14}}{\overset{\kappa_{13}}{\rightleftarrows}} ES_2 \xrightarrow{\kappa_{15}} S_3 + E \\ &S_3 + F \underset{\kappa_{17}}{\overset{\kappa_{16}}{\rightleftarrows}} FS_3 \xrightarrow{\kappa_{18}} S_2 + F \underset{\kappa_{11}}{\overset{\kappa_{10}}{\rightleftarrows}} FS_2 \xrightarrow{\kappa_{12}} S_1 + F \underset{\kappa_5}{\overset{\kappa_4}{\rightleftarrows}} FS_1 \xrightarrow{\kappa_6} S_0 + F, \end{aligned} \tag{4.4.1}$$

where S is the substrate with $n > 1$ phosphorylation sites, S_i denotes the phosphoforms of S with k phosphorylated sites, E and F denote the kinase and phosphatase enzymes as before. There are 12 species: $x_1 = [E], x_2 = [F]$, $x_3 = [S_0]$, $x_4 = [S_1]$, $x_5 = [S_2]$, $x_6 = [S_3]$, $x_7 = [ES_0]$, $x_8 = [FS_1]$, $x_9 = [ES_1]$, $x_{10} = [FS_2], x_{11} = [ES_2], x_{12} = [FS_3]$, and 18 parameters which we denote with the vector $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{18}) \in \mathbb{R}^{18}_{>0}$. The stoichiometric matrix is

$$N = \begin{bmatrix} -1 & 1 & 1 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & 0 & -1 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix},$$

which is of rank 9, and we write a row reduced constraint matrix W whose rows form a basis of $\mathrm{im}(N)^{\perp}$

$$W = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{4.4.2}$$

Hence, the dynamics take place in stoichiometric compatibility classes which are defined by the equations

$$\begin{aligned} x_1 + x_7 + x_9 + x_{11} = c_1, \quad x_2 + x_8 + x_{10} + x_{12} = c_2, \\ x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} = c_3, \end{aligned}$$

for a given $\boldsymbol{c} \in \mathbb{R}^3$. We remove the redundant equations that define the steady states, and write the map described in (4.1.9) for 3-site phosphorylation:

$$\varphi_c(\boldsymbol{x}) = \begin{bmatrix} x_1 + x_7 + x_9 + x_{11} - c_1 \\ x_2 + x_8 + x_{10} + x_{12} - c_2 \\ x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} - c_3 \\ -k_4 x_2 x_4 - k_7 x_1 x_4 + k_3 x_7 + k_5 x_8 + k_8 x_9 + k_{12} x_{10} \\ -k_{10} x_2 x_5 - k_{13} x_1 x_5 + k_9 x_9 + k_{11} x_{10} + k_{14} x_{11} + k_{18} x_{12} \\ -k_{16} x_2 x_6 + k_{15} x_{11} + k_{17} x_{12} \\ k_1 x_1 x_3 - k_2 x_7 - k_3 x_7 \\ k_4 x_2 x_4 - k_5 x_8 - k_6 x_8 \\ k_7 x_1 x_4 - k_8 x_9 - k_9 x_9 \\ k_{10} x_2 x_5 - k_{11} x_{10} - k_{12} x_{10} \\ k_{13} x_1 x_5 - k_{14} x_{11} - k_{15} x_{11} \\ k_{16} x_2 x_6 - k_{17} x_{12} - k_{18} x_{12} \end{bmatrix}. \tag{4.4.3}$$

The Michaelis-Menten constants of each phosphorylation/dephosphorylation event is given as

$$K_1 = \tfrac{\kappa_2+\kappa_3}{\kappa_1},\ K_2 = \tfrac{\kappa_5+\kappa_6}{\kappa_4},\ K_3 = \tfrac{\kappa_8+\kappa_9}{\kappa_7},\ K_4 = \tfrac{\kappa_{11}+\kappa_{12}}{\kappa_{10}},\ K_5 = \tfrac{\kappa_{14}+\kappa_{15}}{\kappa_{13}},\ K_6 = \tfrac{\kappa_{17}+\kappa_{18}}{\kappa_{16}}, \tag{4.4.4}$$

and in what follows we focus on the set of parameters $\boldsymbol{\eta} = (K_1, \ldots, K_6, \kappa_3, \kappa_6, \ldots, \kappa_{18}) \in \mathbb{R}^{15}_{>0}$ instead of $\boldsymbol{\kappa}$. By solving the equations $f_i = 0$ for $i = 4, \ldots, 12$, we get a positive parameterization of the steady states $\Phi(x_1, x_2, x_3) : \mathbb{R}^3_{>0} \to V \cap \mathbb{R}^{12}_{>0}$:

$$\begin{aligned} \Phi_{\boldsymbol{\eta}}(x_1, x_2, x_3) = \Bigg(& x_1, x_2, x_3, \frac{K_2 k_3 x_1 x_3}{k_6 K_1 x_2}, \frac{K_4 x_3 {x_1}^2 k_9 k_3 K_2}{{x_2}^2 K_1 K_3 k_{12} k_6}, \frac{K_6 k_{15} x_3 {x_1}^3 k_9 k_3 K_2 K_4}{k_{18} k_{12} {x_2}^3 k_6 K_1 K_3 K_5}, \frac{x_3 x_1}{K_1}, \\ & \frac{k_3 x_1 x_3}{k_6 K_1}, \frac{x_3 {x_1}^2 k_3 K_2}{x_2 K_1 K_3 k_6}, \frac{x_3 {x_1}^2 k_9 k_3 K_2}{x_2 K_1 K_3 k_{12} k_6} \frac{x_3 {x_1}^3 k_9 k_3 K_2 K_4}{k_{12} {x_2}^2 k_6 K_1 K_3 K_5}, \frac{k_{15} x_3 {x_1}^3 k_9 k_3 K_2 K_4}{k_{18} k_{12} {x_2}^2 k_6 K_1 K_3 K_5} \Bigg). \end{aligned} \tag{4.4.5}$$

As before, we denote the Jacobian of the map $\varphi_c(\boldsymbol{x})$ with $M(\boldsymbol{x})$, and consider the determinant of $M(\Phi(x_1, x_2, x_3))$. The denominator of $\det M(\Phi(x_1, x_2, x_3))$ is positive for any $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$, and the numerator is:

$$
\begin{aligned}
p_{\boldsymbol{\eta}}^{(3)}(\boldsymbol{x}) \;=\; & -K_1K_2K_3^2K_5k_3k_6k_{12}^2k_{18}^2(3K_1K_4k_6k_9 - K_1K_5k_6k_9 + K_2K_5k_3k_{12} + K_3K_5k_3k_{12} - K_4K_5k_6k_9)\; x_1^2x_2^5x_3 \\
& + K_1K_2K_3K_4K_5k_3^2k_6k_9k_{12}k_{18}(K_2K_3k_{12}k_{15} - K_2K_4k_9k_{18} - K_2K_5k_9k_{18} + K_2K_6k_{12}k_{15} - 3K_3K_6k_{12}k_{15})\; x_1^4x_2^3x_3 \\
& + 2K_1K_2K_3^2K_4K_5k_3k_6k_9k_{12}^2k_{18}(K_1k_6k_{15} - K_2k_3k_{18} - K_5k_3k_{18} + K_6k_6k_{15})\; x_1^3x_2^4x_3 \\
& - K_1K_2K_3^2K_5k_3k_6k_{12}^2k_{18}(2K_4k_3k_9k_{18} - 2K_4k_6k_9k_{15} + K_5k_3k_{12}k_{18} - K_5k_6k_9k_{18})\; x_1^3x_2^5x_3 \\
& - K_2^2K_3K_4K_5k_3^2k_9k_{12}k_{18}(K_1k_6k_9k_{18} - K_1k_6k_{12}k_{15} + 2K_3k_3k_{12}k_{18} - 2K_3k_6k_{12}k_{15} + K_5k_3k_{12}k_{18} - K_5k_6k_9k_{18})\; x_1^4x_2^3x_3^2 \\
& - K_2^2K_4^2k_3^2k_9^2(K_2K_5k_3k_9k_{18}^2 - K_2K_5k_3k_{12}k_{15}k_{18} + 2K_3K_6k_3k_{12}k_{15}k_{18} - 2K_3K_6k_6k_{12}k_{15}^2)\; x_1^6x_2x_3^2 \\
& - K_2^2K_3K_4K_5k_3^2k_9k_{12}k_{18}(K_2k_3k_9k_{18} - K_2k_3k_{12}k_{15} + 2K_4k_3k_9k_{18} - 2K_4k_6k_9k_{15} + K_6k_3k_{12}k_{15} - K_6k_6k_9k_{15})\; x_1^5x_2^2x_3^2 \\
& - K_2K_3^2K_5k_3k_{12}^2k_{18}(2K_1K_4k_3k_6k_9k_{18} - 2K_1K_4k_6^2k_9k_{15} + K_2K_5k_3^2k_{12}k_{18} - K_2K_5k_3k_6k_9k_{18})\; x_1^3x_2^4x_3^2 \\
& - K_1K_2K_3K_4K_5k_3k_6k_9k_{12}k_{18}(K_2k_3k_9k_{18} - K_2k_3k_{12}k_{15} + 2K_3k_3k_{12}k_{18} - 2K_3k_6k_{12}k_{15})\; x_1^4x_2^4x_3 \qquad (4.4.6) \\
& - K_1K_2K_3^2K_5^2k_3k_6k_{12}^2k_{18}^2(k_3k_{12} - k_6k_9)\; x_3x_2^6x_1^2 - K_1K_2K_3^2K_5^2k_3k_6k_{12}^2k_{18}^2(k_3k_{12} - k_6k_9)\; x_1^2x_2^5x_3^2 \\
& - K_2^3K_4^2K_6k_3^3k_9^2k_{15}(k_9k_{18} - k_{12}k_{15})\; x_3^2x_1^7 - K_1K_2^2K_3K_4K_5k_3^2k_6k_9k_{12}k_{18}(k_9k_{18} - k_{12}k_{15})\; x_1^5x_2^3x_3 \\
& - 2K_2K_5^2k_3k_{12}^3k_{18}^2k_6^2K_1^2K_3^2\; x_1x_2^6x_3 - 2K_2^2k_3^2k_9^2K_6k_{15}K_4k_{18}k_{12}k_6K_1K_3K_5\; x_1^5x_2^2x_3 \\
& - k_{18}^2k_{12}^3k_6^3K_1^3K_3^3K_5^2\; x_2^7 - K_1^2K_3^3K_5^2k_6^2k_{12}^3k_{18}^2(k_3+k_6)\; x_1x_2^7 - K_3^3K_5^2k_6^3k_{12}^3k_{18}^2K_1^2\; x_2^7x_3 \\
& - K_2K_4K_6k_3k_9k_{15}k_{18}k_{12}^2k_6^2K_1^2K_3^2K_5\; x_1^3x_2^4 - K_2K_4K_5^2k_3k_9k_{18}^2k_{12}^2k_6^2K_1^2K_3^2\; x_1^2x_2^5 \\
& - K_1^2K_2K_3^2K_4K_5k_3k_6^2k_9k_{12}^2k_{18}(k_{15}+k_{18})\; x_1^3x_2^5 - K_2K_3^3K_5^2k_3k_{12}^3k_{18}^2k_6^2K_1^2\; x_1x_2^6 \\
& - K_1^2K_2K_3^2K_5^2k_3k_6^2k_{12}^2k_{18}^2(k_9+k_{12})\; x_1^2x_2^6 - K_2^2K_4^2k_3^2k_9^2k_{15}^2K_6k_{12}x_1^6k_6K_1K_3\; x_1^6x_2x_3.
\end{aligned}
$$

See the supplementary file *SupplementaryInfoThesis.mw* for the full steps in the calculation of $p_\kappa^{(3)}(\boldsymbol{x})$. In order to study the multistationarity of the 3-site network, we consider the following analog of Proposition 4.2.4.

Proposition 4.4.1 (Analog of Proposition 4.2.4)**.** Let $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$ be fixed, and $p_{\boldsymbol{\eta}}^{(3)}(\boldsymbol{x})$ be given as in (4.4.6). Then, the following statements hold.

(Mono) If $p_{\boldsymbol{\eta}}^{(3)}(\boldsymbol{x})$ is negative for all $x_1, x_2, x_3 > 0$, then $\boldsymbol{\eta}$ does not enable multistationarity, and there is exactly one positive steady state in each invariant linear subspace.

(Mult) If $p_{\boldsymbol{\eta}}^{(3)}(\boldsymbol{x})$ is positive for some $x_1, x_2, x_3 > 0$, then $\boldsymbol{\eta}$ enables multistationarity, in the invariant linear subspace containing the point $\Phi_{\boldsymbol{\eta}}(x_1, x_2, x_3)$, where $\Phi_{\boldsymbol{\eta}}$ is given as in (4.2.10).

Proof. By following the exact arguments from the proof of Proposition 4.2.4, we see that the network in 3-site is dissipative, and it does not have any boundary steady states. Moreover, we point out a positive parameterization of the steady states in (4.2.10). Therefore we can invoke Theorem 4.1.12. Since the rank of the system is 9 in 3-site case, the sign conditions in the statement are flipped. □

Note that, given $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$, $p_{\boldsymbol{\eta}}^{(3)}(\boldsymbol{x}) < 0$ for all $x_1, x_2, x_3 > 0$ if and only if $-p_{\boldsymbol{\eta}}^{(3)}(\boldsymbol{x}) > 0$ for all $x_1, x_2, x_3 > 0$. Therefore, if any necessary condition on $\boldsymbol{\eta}$ certifying the nonnegativity of $-p_{\boldsymbol{\eta}}^{(3)}(\boldsymbol{x})$ yields a condition for $\boldsymbol{\eta}$ to enable monostationarity. For the convenience

of notation let $q_{\boldsymbol{\eta}}(\boldsymbol{x}) := -p^{(3)}_{\boldsymbol{\eta}}(\boldsymbol{x})$. The vertices of New $(q_{\boldsymbol{\eta}}(\boldsymbol{x}))$ are

$$(2,6,1),(2,5,2),(5,3,1),(7,0,2),(6,1,1),(0,7,0),(1,7,0),(0,7,1),(3,4,0),(3,5,0).$$

The support of $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ clearly contains more exponents than the polynomial arising from the 2-site case. However, one can verify, for example the function via `isomorphic` in `POLYMAKE`, that the Newton polytopes of the polynomials given in (4.2.12) and (4.4.6) have isomorphic face lattices. We observe that 11 out of 24 terms of $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ have positive coefficients for every $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$. We note that the coefficients of $(2,6,1)$ and $(2,5,2)$ contains the factor $\kappa_3\kappa_{12} - \kappa_6\kappa_9$, and the coefficients of $(7,4,1)$ and $(10,0,2)$ contains the factor $\kappa_9\kappa_{18} - \kappa_{12}\kappa_{15}$. The coefficients of the rest of the vertices are positive for $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$. Hence, if $\kappa_3\kappa_{12} - \kappa_6\kappa_9 < 0$ or $\kappa_9\kappa_{18} - \kappa_{12}\kappa_{15} < 0$, then $\boldsymbol{\eta}$ enables multistationarity. Therefore, in what follows we assume that

$$\kappa_3\kappa_{12} - \kappa_6\kappa_9 > 0 \qquad \kappa_9\kappa_{18} - \kappa_{12}\kappa_{15} > 0. \tag{4.4.7}$$

Under the conditions given in (4.4.7), the coefficients of the terms corresponding to the following six exponents of $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ become positive:

$$(3,5,1),(4,3,2),(6,1,2),(7,0,2),(5,2,2),(3,4,2),(4,4,1).$$

As an example, consider the coefficient of the term corresponding to the exponent $(3,5,1)$. This coefficient is nonnegative if and only if

$$A := 2K_4k_3k_9k_{18} - 2K_4k_6k_9k_{15} + K_5k_3k_{12}k_{18} - K_5k_6k_9k_{18} > 0.$$

If we impose the conditions on (4.4.7), then we have

$$A > 2K_4k_3k_9k_{18} - 2K_4k_3k_9k_{18} + K_5k_3k_{12}k_{18} - K_5k_3k_{12}k_{18} = 0.$$

Using the same argument, we see that the coefficient of the remaining five points also become positive under the assumptions (4.4.7). Therefore, under the assumptions (4.4.7) the only terms that can have negative coefficients are the ones corresponding to the following exponents:

$$\boldsymbol{c_1} = (2,5,1), \quad \boldsymbol{c_2} = (3,4,1), \quad \boldsymbol{c_3} = (4,3,1). \tag{4.4.8}$$

Note that $\boldsymbol{c_1}, \boldsymbol{c_2}$ and, $\boldsymbol{c_3}$ are contained in a single face of the New $(q_{\boldsymbol{\eta}}(\boldsymbol{x}))$, which is

given as $H := \text{conv}(\{\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}, \boldsymbol{a_4}, \boldsymbol{a_5}, \boldsymbol{a_6}\})$ where

$$\boldsymbol{a_1} = (0,7,1),\ \boldsymbol{a_2} = (2,5,2),\ \boldsymbol{a_3} = (7,0,2),\ \boldsymbol{a_4} = (6,1,1),\ \boldsymbol{a_5} = (3,4,0),\ \boldsymbol{a_6} = (0,7,0). \tag{4.4.9}$$

Let $q_{\boldsymbol{\eta},H}(\boldsymbol{x})$ denote the restriction of $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ to the face H of its Newton polytope. Then, due to Proposition 4.2.7, $q_{\boldsymbol{\eta},H}(\boldsymbol{x})$ is nonnegative if and only if $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ is. Based on this fact, we present an analog of Theorem 4.3.5 for 3-site phosphorylation. We set some notation for three values that depend on $\boldsymbol{\eta}$, and we define

$$\begin{aligned}
C_1(\boldsymbol{\eta}) &:= \Big(54K_2{K_3}^2K_5\kappa_3{\kappa_{12}}^2\kappa_{18}\big(2K_1K_4\kappa_3\kappa_6\kappa_9\kappa_{18} - 2K_1K_4{\kappa_6}^2\kappa_9\kappa_{15} + K_2K_5{\kappa_3}^2\kappa_{12}\kappa_{18} - K_2K_5\kappa_3\kappa_6\kappa_9\kappa_{18}\big)\Big)^{\frac{1}{3}} \\
&\quad \left(K_2K_4{K_5}^2\kappa_3\kappa_9{\kappa_{18}}^2{\kappa_6}^2{K_1}^2{K_3}^2{\kappa_{12}}^2\right)^{\frac{1}{3}} \left(K_2{K_5}^2\kappa_3{\kappa_{12}}^3{\kappa_{18}}^2{\kappa_6}^2{K_1}^2{K_3}^2\right)^{\frac{1}{3}} \\
C_2(\boldsymbol{\eta}) &:= 2\left({K_3}^3{K_5}^2{\kappa_6}^3{\kappa_{12}}^3{\kappa_{18}}^2{K_1}^2\right)^{\frac{1}{2}} \left({K_2}^2{K_4}^2{\kappa_3}^2{\kappa_9}^2{\kappa_{15}}^2K_6\kappa_6K_1K_3\kappa_{12}\right)^{\frac{1}{2}} \\
C_3(\boldsymbol{\eta}) &:= \left(54K_2K_4K_6\kappa_3\kappa_9\kappa_{15}{\kappa_6}^2{K_1}^2{K_3}^2{\kappa_{12}}^2K_5\kappa_{18}\right)^{\frac{1}{3}} \Big({K_2}^2K_3K_4K_5{\kappa_3}^2\kappa_9\kappa_{12}\kappa_{18}\big(K_1\kappa_6\kappa_9\kappa_{18} - K_1\kappa_6\kappa_{12}\kappa_{15} \\
&\quad + 2K_3\kappa_3\kappa_{12}\kappa_{18} - 2\,K_3\kappa_6\kappa_{12}\kappa_{15} + K_5\kappa_3\kappa_{12}\kappa_{18} - K_5\kappa_6\kappa_9\kappa_{18}\big)\Big)^{\frac{1}{3}} \left({K_2}^2{\kappa_3}^2{\kappa_9}^2K_6\kappa_{15}K_4\kappa_6K_1K_3\kappa_{12}K_5\kappa_{18}\right)^{\frac{1}{3}}.
\end{aligned} \tag{4.4.10}$$

Theorem 4.4.2 (Analog of Theorem 4.3.5)**.** Let $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$ be a vector that satisfies the conditions in (4.4.7), and $C_1(\boldsymbol{\eta}), C_2(\boldsymbol{\eta})$ and $C_3(\boldsymbol{\eta})$ be given as in (4.4.10). For $\boldsymbol{c_1}, \boldsymbol{c_2}$ and, $\boldsymbol{c_3}$ given as in (4.4.8). Furthermore, let $c_{\boldsymbol{\eta},\boldsymbol{c_1}}$, $c_{\boldsymbol{\eta},\boldsymbol{c_2}}$ and $c_{\boldsymbol{\eta},\boldsymbol{c_3}}$ denote the coefficients of terms $\boldsymbol{x}^{\boldsymbol{c_1}}$, $\boldsymbol{x}^{\boldsymbol{c_2}}$ and $\boldsymbol{x}^{\boldsymbol{c_3}}$ in the polynomial $q_{\boldsymbol{\eta}}$, respectively. If for some $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$

(i) $C_1(\boldsymbol{\eta}) \geq -c_{\boldsymbol{\eta},\boldsymbol{c_1}}$,

(ii) $C_2(\boldsymbol{\eta}) \geq -c_{\boldsymbol{\eta},\boldsymbol{c_2}}$,

(iii) $C_3(\boldsymbol{\eta}) \geq -c_{\boldsymbol{\eta},\boldsymbol{c_3}}$,

hold simultaneously, then $q_{\boldsymbol{\eta}}$ is nonnegative over $\mathbb{R}^3_{>0}$. Consequently, $\boldsymbol{\eta}$ precludes multistationarity.

Proof. Let $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$ be a parameter vector such that the conditions (i), (ii) and (iii) in the statement hold. First, we show that $q_{\boldsymbol{\eta},H}$ is nonnegative. Consider the Newton polytope of $q_{\boldsymbol{\eta},H}$ which is depicted in Figure 4.12. Besides the exponent points $\boldsymbol{a_1}, \ldots, \boldsymbol{a_6}, \boldsymbol{c_1}, \boldsymbol{c_2}, \boldsymbol{c_3}$, there are 8 more exponents of $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ in the face H of New $(q_{\boldsymbol{\eta}}(\boldsymbol{x}))$. We denote these points

as follows:

$$\begin{aligned} \boldsymbol{b_1} &= (3,4,2), & \boldsymbol{b_2} &= (4,3,2), \\ \boldsymbol{b_3} &= (5,2,2), & \boldsymbol{b_4} &= (6,1,2), \\ \boldsymbol{b_5} &= (2,5,0), & \boldsymbol{b_6} &= (1,7,0), \\ \boldsymbol{i_1} &= (1,6,1), & \boldsymbol{i_2} &= (5,2,1). \end{aligned} \tag{4.4.11}$$

Only the coefficients of $\boldsymbol{c_1}, \boldsymbol{c_2}$ and $\boldsymbol{c_3}$ can be negative in the polynomial $q_{\boldsymbol{\eta}}(\boldsymbol{x})$. Following the proof of Theorem 4.3.5, we write three circuit polynomials, one for each $\boldsymbol{c_i}$.

Remember that given an exponent $\boldsymbol{\alpha}$ of $q_{\boldsymbol{\eta}}(\boldsymbol{x})$, we denote its coefficient in $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ by $c_{\boldsymbol{\eta},\boldsymbol{\alpha}}$. We define a circuit polynomial $q_{\boldsymbol{\eta},1}$ whose support is given by the 2-dimensional simplex $\boldsymbol{b_1}, \boldsymbol{i_2}, \boldsymbol{b_5}$ and the inner term exponent $\boldsymbol{c_1}$ as follows:

$$q_{\boldsymbol{\eta},1}(x_1,x_2,x_3) = c_{\boldsymbol{\eta},\boldsymbol{b_1}}\boldsymbol{x}^{\boldsymbol{b_1}} + c_{\boldsymbol{\eta},\boldsymbol{i_2}}x^{\boldsymbol{i_2}} + c_{\boldsymbol{\eta},\boldsymbol{b_5}}x^{\boldsymbol{b_5}} + \bar{c}_{\boldsymbol{\eta},1}\boldsymbol{x}^{\boldsymbol{c_1}},$$

Note that $c_{\boldsymbol{\eta},\boldsymbol{b_1}}$, $c_{\boldsymbol{\eta},\boldsymbol{i_2}}$, $c_{\boldsymbol{\eta},\boldsymbol{b_5}}$ are equal to the coefficients of $\boldsymbol{x}^{\boldsymbol{b_1}}$,$\boldsymbol{x}^{\boldsymbol{i_2}}$,$\boldsymbol{x}^{\boldsymbol{b_5}}$ in $q_{\boldsymbol{\eta},H}(\boldsymbol{x})$, respectively, and $\bar{c}_{\boldsymbol{\eta},1}$ is in $\mathbb{R}$. In a similar fashion, we define the polynomials $q_{\boldsymbol{\eta},2}(x_1,x_2,x_3)$ and $q_{\boldsymbol{\eta},3}(x_1,x_2,x_3)$, whose supports are respectively given by $\boldsymbol{c_2}$ as inner term with 1-dimensional simplex $\boldsymbol{a_1}, \boldsymbol{a_4}$, and $\boldsymbol{c_3}$ as inner term with 2-dimensional simplex $\boldsymbol{a_5}, \boldsymbol{b_2}, \boldsymbol{i_2}$. As before, we let $\bar{c}_{\boldsymbol{\eta},i} \in \mathbb{R}$ be the coefficient of $\boldsymbol{x}^{\boldsymbol{c_i}}$ in the respective polynomial $q_{\boldsymbol{\eta},i}$. Furthermore, the coefficients of remaining terms in each $q_{\boldsymbol{\eta},i}$ is assumed to be equal to the coefficient of the same term in $q_{\boldsymbol{\eta},H}$. We compute the circuit numbers of each $q_{\boldsymbol{\eta},i}$:

$$\Theta_{q_{\boldsymbol{\eta},1}} = 3\left(c_{\boldsymbol{\eta},\boldsymbol{b_1}}c_{\boldsymbol{\eta},\boldsymbol{i_1}}c_{\boldsymbol{\eta},\boldsymbol{b_5}}\right)^{\frac{1}{3}}, \quad \Theta_{q_{\boldsymbol{\eta},2}} = 2\left(c_{\boldsymbol{\eta},\boldsymbol{a_1}}c_{\boldsymbol{\eta},\boldsymbol{a_4}}\right)^{\frac{1}{2}}, \quad \Theta_{q_{\boldsymbol{\eta},3}} = 3\left(c_{\boldsymbol{\eta},\boldsymbol{a_5}}c_{\boldsymbol{\eta},\boldsymbol{i_2}}c_{\boldsymbol{\eta},\boldsymbol{b_5}}\right)^{\frac{1}{3}}$$

Given a parameter vector $\boldsymbol{\eta} = (K_1,\ldots,K_6,\kappa_3,\kappa_6,\ldots,\kappa_{18})$, the circuit numbers $\Theta_{q_{\boldsymbol{\eta},1}}$, $\Theta_{q_{\boldsymbol{\eta},2}}$ and $\Theta_{q_{\boldsymbol{\eta},3}}$ are equal to $C_1(\boldsymbol{\eta})$, $C_2(\boldsymbol{\eta})$ and $C_3(\boldsymbol{\eta})$ in (4.4.10), respectively. Therefore, if the statements (i),(ii) and (iii) hold simultaneously, then Corollary 4.2.11 implies that each $q_{\boldsymbol{\eta},i}$ is nonnegative over $\mathbb{R}^3_{>0}$. Consequently, $q_{\boldsymbol{\eta}}$ is nonnegative, and due to Proposition 4.4.1 $\boldsymbol{\eta}$ the precludes multistationarity. □

Example 4.4.3. Fix $K_2 = 1$, $K_3 = 1$, $K_5 = 1$, $K_6 = 1$, $k_3 = 2$, $k_6 = 1$, $k_9 = 1$, $k_{12} = 1$, $k_{15} = 1$, $k_{18} = 2$, and consider the set of parameters $\boldsymbol{\eta} = (K_1,1,1,K_4,1,1,2,1,1,1,1,2)$

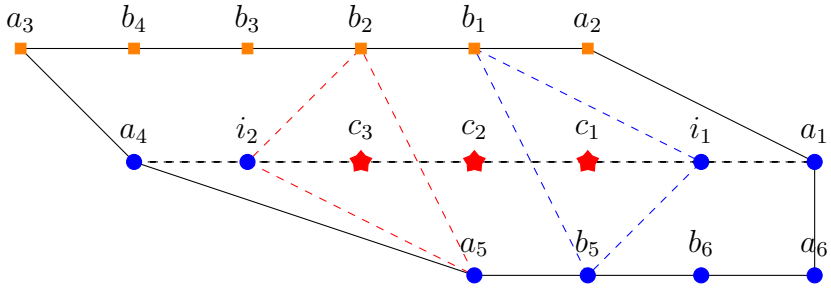

Figure 4.12: The figure depicts the Newton polytope of $q_{\boldsymbol{\eta},H}(\boldsymbol{x})$, where each blue dot is an exponent whose coefficient is positive for all $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$, each orange squares is an exponent whose coefficient is positive if $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$ satisfies (4.4.7), and each red star is an exponent whose coefficient can be negative even if $\boldsymbol{\eta} \in \mathbb{R}^{15}_{>0}$ satisfies (4.4.7). The coordinates of the vertices $\boldsymbol{a_1}, \ldots, \boldsymbol{a_6}$ are given as in (4.4.9), the points $\boldsymbol{b_1}, \cdot, \boldsymbol{b_6}, \boldsymbol{i_1}$ and $\boldsymbol{i_2}$ are given as in (4.4.11), and the points $\boldsymbol{c_1}, \boldsymbol{c_2}$ and $\boldsymbol{c_3}$ are given as in (4.4.8). The dashed blue triangle (on right) is the Newton polytope of $q_{\boldsymbol{\eta},1}$, the dashed line segment is the Newton polytope of $q_{\boldsymbol{\eta},2}$, and the dashed red triangle (on left) is the Newton polytope of $q_{\boldsymbol{\eta},3}$ from the proof of Theorem 4.4.2.

for $K_1, K_4 > 0$. Then the inequalities given in the statement of Theorem 4.4.2 reduces to:

$$\begin{aligned}
&\text{(i) } 12\left(K_4{K_1}^2\right)^{\frac{1}{3}}\left(-8K_4\left(-8-K_1\right)\right)^{\frac{1}{3}}\left(K_4K_1\right)^{\frac{1}{3}} \geq 8K_4K_1 + 4\left(-4{K_4}^2 - 4K_4\right)K_1,\\
&\text{(ii) } 8K_1\sqrt{K_4^2K_1} \geq 32K_4K_1 + \left(24K_1 - 8{K_1}^2\right)K_4, \qquad (4.4.12)\\
&\text{(iii) } 12\left(K_4{K_1}^2\right)^{\frac{1}{3}}\left(-8K_4\left(-8-K_1\right)\right)^{\frac{1}{3}}\left(K_4K_1\right)^{\frac{1}{3}} \geq 8K_4K_1 + 4\left(4{K_4}^2 + 4K_4\right)K_1.
\end{aligned}$$

Each inequality above cuts a region in the parameter space. By intersecting all three regions, we can find a region in the parameter space where we can certify that $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ is nonnegative. We note that this intersection is nonempty. In particular, consider the point $\boldsymbol{\eta} = (7,1,1,1,1,1,2,1,1,1,1,2)$ in the parameter space, and put $K_1 = 7$ and $K_4 = 1$ in for each inequality (i), (ii) and (iii) above. The left hand sides are evaluated to $\approx 1151.5, \approx 148.2, \approx 414.3$, and right hand sides are evaluated to -952, 0, 280, respectively. Since inequalities (i), (ii) and (iii) are satisfied simultaneously for $\boldsymbol{\eta} = (7,1,1,1,1,1,2,1,1,1,1,2)$, it precludes multistationarity. We also verify the fact that $\boldsymbol{\eta} = (7,1,1,1,1,1,2,1,1,1,1,2)$ precludes multistationarity by computing $q_{\boldsymbol{\eta}}(\boldsymbol{x})$ for this specific value of $\boldsymbol{\eta}$:

$$\begin{aligned}
q_{\boldsymbol{\eta}}(\boldsymbol{x}) = {} & 8{x_3}^2{x_1}^7 + 40x_2{x_3}^2{x_1}^6 + 56{x_2}^3x_3{x_1}^5 + 72{x_2}^2{x_3}^2{x_1}^5 + 224{x_2}^4x_3{x_1}^4 + 120{x_2}^3{x_3}^2{x_1}^4\\
& + 224x_3{x_2}^5{x_1}^3 + 184{x_2}^4{x_3}^2{x_1}^3 + 56x_3{x_2}^6{x_1}^2 + 56{x_3}^2{x_2}^5x_1^2 + 28{x_1}^6x_3x_2 + 112{x_1}^5x_3{x_2}^2\\
& + 280{x_2}^3x_3{x_1}^4 + 588x_2^5{x_1}^3 + 784{x_2}^6{x_1}^2 + 952x_3{x_2}^5{x_1}^2 + 588{x_2}^7x_1 + 784x_1{x_2}^6x_3\\
& + 196{x_2}^7x_3 + 196{x_1}^3{x_2}^4 + 392{x_1}^2{x_2}^5 + 392x_1{x_2}^6 + 1372{x_2}^7,
\end{aligned}$$

which is clearly nonnegative for $\boldsymbol{x} \in \mathbb{R}^3$. See Figure 4.13 for a plot of the curves that arise from the conditions (i), (ii) and (iii). ⬡

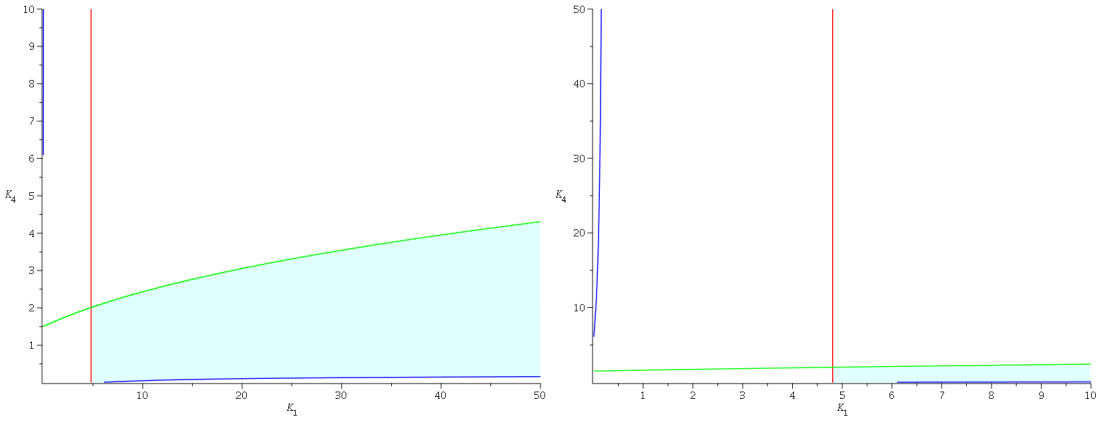

Figure 4.13: With $(K_2, K_3, K_5, K_6, \kappa_3, \kappa_6, \ldots, \kappa_{18}) = (1, 1, 1, 1, 2, 1, 1, 1, 1, 2)$, the figure shows three curves given by the inequalities in (4.4.12). Blue curve corresponds to (i) from (4.4.12), red curve corresponds to (ii) from (4.4.12), green curve corresponds to (iii) from (4.4.12).

To conclude this section, we briefly discuss how to extend the approach that is used to prove Theorem 4.4.2 to the phosphorylation cycles with higher number of sites. Our preliminary study of the Newton polytope for 4-site and shows that, the face lattice of the Newton polytope in 4-site is isomorphic to the face lattices in the 2-site and 3-site cases. Indeed, the Newton polytope is given by the 10 vertices

$$(2, 9, 1), (2, 8, 2), (7, 4, 1), (10, 0, 2), (8, 2, 1), (4, 7, 0), (4, 6, 0), (1, 10, 0), (0, 10, 1), (0, 10, 0)$$

in the 4-site case. Since the expressions become too long to be human readable, we refer to the supplementary file *SupplementaryInfoThesis.mw* for details on calculating the vertices. Then, using the function `isomorphic` in `POLYMAKE`, we can verify that the polytope given by the vertices above is combinatorially equivalent to the Newton polytopes $\text{New}\left(p_{\boldsymbol{\eta}}^{(3)}\right)$ from (4.4.6) and $\text{New}\left(p_{\boldsymbol{\eta}}(x)\right)$ from (4.2.12). Thus, we make the following conjecture.

Conjecture 4.4.4. Given any $n \in \mathbb{N}$ such that $n \geq 2$, let p be the polynomial that arise as the determinant in Theorem 4.1.12 from the n-site phosphorylation cycle. Then, the face lattice of $\text{New}\,(p)$ is isomorphic to the face lattice of $\text{New}\left(p_{\boldsymbol{\eta}}(x)\right)$ from (4.2.12).

As the number of sites increases, the number of the exponent points that can have negative coefficient increases too. This implies that, if we want to use an approach similar to Theorem 4.4.2, then the number of inequalities, that one has to check, increases. However, our initial experiments verified that, the exponent points with negative coefficients can only be appear in certain facets of Newton polytope.

Conjecture 4.4.5. Given any $n \in \mathbb{N}$ such that $n \geq 2$, let p be the polynomial that arise as the determinant in Theorem 4.1.12 from the n-site phosphorylation cycle. Let $\boldsymbol{\alpha} \in A_p$ be an exponent of p so that the coefficient of p at $\boldsymbol{\alpha}$ may take negative values for some parameter vector $\boldsymbol{\kappa}$. Then, $\boldsymbol{\alpha}$ either belongs to the hexagonal face of $\mathrm{New}\,(p)$, or to a 1-dimensional face. Furthermore, every exponent of p which does not lie on the hexagonal face with a possibly negative coefficient, must lie on the same 1-dimensional face.

Chapter 5

Resume

In this final section, we revisit the investigated problems throughout this thesis, recall our contributions and remark the problems that remain open. We address to these in under two main titles.

Foundations of Maximal Mediated Sets

In Chapter 3, we study the maximal mediated sets associated to the integral simplices Δ with $\mathrm{Vert}\,(\Delta) \subset (2\mathbb{Z})^n$. Section 3.1 mostly consists of definitions and facts given prior to this work. However, there are some novelties that worth of mentioning. First, we observe that the union of two Δ-mediated sets is also Δ-mediated in Proposition 3.1.6, and point out Corollary 3.1.8 based on this observation. This corollary leads to a novel proof of the existence and the uniqueness of the maximal mediated sets, which we present in Theorem 3.1.11. We use Algorithm 3.1.14 to compute the maximal mediated sets, which is a different approach than Algorithm 3.1.12 that was used by Reznick in his original proof of Theorem 3.1.11. We present a full proof of Algorithm 3.1.14 in this thesis, and we note that the specific implementation that we use for computing maximal mediated sets was written by Olivia Röhrig on `POLYMAKE` as a part of [HRdWY20] and her master thesis [Roe20]. Moreover, we prove Theorem 3.1.26 in Section 3.1.2, which does not only yield Corollary 3.1.27, but also generalizes Theorem 3.1.1. As a consequence of this generalization, we see that the question of whether a SONC polynomial with simplex Newton polytope is a sum of squares is closely related to the maximal mediated set associated to its Newton polytope.

In Section 3.2, we start by defining the notion of the h-ratio (Definition 3.2.1) to measure the density of Δ^* in $\mathrm{conv}\,\Delta \cap \mathbb{Z}^n$, and study the affine transformations that preserve the h-ratio (Definition 3.2.3). In Theorem 3.2.6, we give a classification of transformations that preserve the maximal mediated set structure, in particular the h-ratio. Consequently, we show that the maximal mediated set structure of a simplicial basin is defined by the

underlying lattice described in Corollary 3.2.7.

In Section 3.3, we provide a large database of MMS which was generated in collaboration with Olivia Röhrig using `POLYMAKE`. The database is partitioned according to the number of variables, n, maximal degree of the circuit polynomials $2d$ for $n, 2d$ given in Table 3.1 and a sampled database with a similar partitioning for n and $2d$ given in Table 3.2.

https://polymake.org/downloads/MMS/

Corollary 3.2.7 leads to a process called HNF reduction (Remark 3.3.2), which reduces the size of the database substantially and smoothens the spikes we observe in the data in exchange of an increased run time. Furthermore, we observe that HNF reduction has a nontrivial effect on the h-ratio distribution. Since HNF reduction changes the mean of the h-ratio distribution, we see that the HNF reduction in fact yields a different distribution. Moreover, for $n = 2$ up to $2d = 150$ we show that:

1. the distribution of h-ratio is a Bernoulli distribution.

2. expected value is close to 1.

The first part computationally proves a conjecture given by Reznick in 1989, Conjecture 3.1.19, up to $2d = 150$. Yet, the full proof of the conjecture is still missing. The second observation follows then from (1) and [IdW16a, Theorem 5.9].

Based on the computed data, we conjecture that for $n > 2$ as $2d$ grows the distribution of h-ratio over lattices does **not** converge to the uniform distribution. Observing the data, we believe that the distribution of the h-ratio for fixed n as $2d$ grows might be related to the Chi-squared distribution, but we do not have hard evidence for this fact. The main issue is improving the database and accessing to the h-ratio distribution for higher n and $2d$. A possible way to discover the distribution for higher n and $2d$ would be to sample the lattices with uniform standard distribution, and we are not aware how this can be done. More importantly, we need a faster way to determine whether L_{Δ_1} and L_{Δ_2} share the same Hermite normal form up to a permutation of columns. This will not only enable us to compute the exact distribution for more cases, but also increase the speed of sampling.

Symbolic SONC Certificates and Multistationarity in CRNT

In Chapter 4, we have studied the parameter region of multistationarity for a relevant biochemical system in detail, namely a dual phosphorylation cycle, by using ideas and techniques from real algebraic geometry. The dual phosphorylation cycle has been the object of many mathematical analyses and yet, several aspects remain unknown. One of

such aspects that we consider in Chapter 4 concerns the characterization of the multistationarity region in the parameter space of dual phosphorylation cycle. In order to detect multistationarity in dual phosphorylation, we use [CFMW17, Corollary 2], which we state in Proposition 4.2.4, and study the sign of the polynomial $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ given in (4.2.12). Before our results, $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ was used in [CM14] to give two rational functions $a(\boldsymbol{\kappa})$ and $b(\boldsymbol{\kappa})$ on the reaction rate parameters $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{12})$ such that:

- the system precludes multistationarity if $a(\boldsymbol{\kappa}) \geq 0$ and $b(\boldsymbol{\kappa}) \geq 0$,
- and the system has admits multiple steady states in some stoichiometric compatibility class if $a(\boldsymbol{\kappa}) < 0$.

In Section 4.3, we address to the case $a(\boldsymbol{\kappa}) \geq 0$ and $b(\boldsymbol{\kappa}) < 0$ and give a complete characterization of the multistationarity region in terms of kinetic parameters for 2-site phosphorylation cycle. With Theorem 4.3.15 in Section 4.3.3,we provide a full parametric description of the mono- and the multistationarity regions, by giving an explicit parametric representation of the boundary between the two regions. Also, in Theorem 4.3.18, we show that the region of monostationarity is a closed connected set, and the region of multistationarity is an open connected set in $\mathbb{R}^{12}_{>0}$. The proof of Theorem 4.3.15 is based crucial observations on theory of discriminants, which we discuss in Section 4.3.1, and on the theory of circuit polynomials, which we discuss in Section 4.3.2. The content of Section 4.3.1 is not directly related to the circuit polynomials, yet included in this thesis in order to give a complete picture of the results in [FKdWY20].

In Section 4.3.2, we describe an open set in the monostationarity region of the 2-site phosphorylation cycle using the nonnegative circuit polynomials. In particular, we write a necessary condition for the nonnegativity of $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ based on circuit numbers in Theorem 4.3.5. As we emphasize in Remark 4.3.7, the nonnegativity of $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ is fully captured by SONC polynomials in the case of 2-site phosphorylation. Moreover, we give a preliminary description of the monostationarity region in Corollary 4.3.10 using Theorem 4.3.5, which plays a significant role in the proof of Theorem 4.3.15. Furthermore, Corollary 4.3.10 also takes part in proving the connectivity of the monostationarity region in Theorem 4.3.18.

In Section 4.4, we extend the circuit polynomial approach which we use for certifying the preclusion of monostationarity in Theorem 4.3.5. First, we give an interpretation of [CFMW17, Corollary 2] in the case of 3-site phosphorylation, and then, we point out to Proposition 4.4.1, which is an analogous result to Proposition 4.2.4. Consequently, we prove Theorem 4.4.2, which describes a subset of monostationarity region given by three inequalities of generalized polynomials. We further see that the subset we describe is not empty by computing an explicit point from this subset. We conclude the discussion with giving two conjectures Conjecture 4.4.4 and Conjecture 4.4.5, which are still a part of

an ongoing research. We expect that the circuit polynomial approach, which we use for 2-site and 3-site cases, is extendable to the cases with higher number of sites with the help of these two conjectures.

In summary, our results in Chapter 4 have advanced the understanding of the set of kinetic parameters yielding multistationarity for 2- and n-site phosphorylation respectively, and in particular, have shown that this set is connected. The full region of multistationarity, in the parameter space also involving the total amounts of kinase, phosphatase, and substrate, is still unknown. Particularly, it remains an open problem whether it is connected. However, since we have shown that the projection of the full region onto the kinetic parameters is connected, it could potentially be used to decide whether the full region is also connected. The techniques used here target the study of the signs of a parametric multivariate polynomial on the positive orthant as a function of the parameters, and hence are not exclusive to the dual phosphorylation cycle. For instance, the allosteric kinase model given in [FSW$^+$16] presents difficulties analogous to those encountered in this thesis.

For any biochemical system for which the multistationarity region in kinetic space can be studied using the methods in [CFMW17], one can show that an analogous statement to Proposition 4.2.4 holds. Consequently, the multistationarity of such systems can potentially be analyzed following the same steps we take in Section 4.3. However, there is a limit on the system's size, as symbolic algebraic methods are computationally demanding. Furthermore, the study of signs also plays a key role when analyzing the stability of steady states or the presence of Hopf bifurcations via the Routh-Hurwitz criterion (see for example [TF20, CMS19]). Therefore, the techniques we introduced and used throughout this chapter are not only useful to study multistationarity, but also can be used to address questions in other directions in CRNT.

Bibliography

[Art27] E. Artin, *Über die zerlegung definiter funktionen in quadrate*, Abhandlungen aus dem mathematischen Seminar der Universität Hamburg, vol. 5, Springer, 1927, pp. 100–115.

[BCR98] J. Bochnak, M. Coste, and M.-F. Roy, *Real algebraic geometry*, Springer, 1998.

[BDG20] F. Bihan, A. Dickenstein, and M. Giaroli, *Lower bounds for positive roots and regions of multistationarity in chemical reaction networks*, Journal of Algebra **542** (2020), 367–411.

[BG99] W. Bruns and J. Gubeladze, *Rectangular simplicial semigroups*, Commutative algebra, algebraic geometry, and computational methods, Springer Singapore (1999), 201–214.

[BKVH07] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, *A tutorial on geometric programming*, Optimization and engineering **8** (2007), no. 1, 67.

[Ble06] G. Blekherman, *There are significantly more nonegative polynomials than sums of squares*, Israel Journal of Mathematics **153** (2006), no. 1, 355–380.

[Bom14] F. Bommel, *Darstellung von Summen von Quadraten als Projektion niedrigdimensionaler Spektraeder*, Master's thesis, Goethe Universität Frankfurt am Main, 2014.

[BPT12] G. Blekherman, P.A. Parrilo, and R.R. Thomas, *Semidefinite optimization and convex algebraic geometry*, Society for Industrial and Applied Mathematics, 2012.

[BV11] S.P. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2011.

[CDM+11] C. Chen, J.H. Davenport, M.M. Moreno, B Xia, and R. Xiao, *Computing with semi-algebraic sets represented by triangular decomposition*, Proceedings of the 36th international symposium on Symbolic and algebraic computation, 2011, pp. 75–82.

[CF12] C. Conradi and D. Flockerzi, *Multistationarity in mass action networks with applications to ERK activation*, Journal of mathematical biology **65** (2012), no. 1, 107–156.

[CFM20] C. Conradi, E. Feliu, and M. Mincheva, *On the existence of Hopf bifurcations in the sequential and distributive double phosphorylation cycle*, Mathematical Biosciences and Engineering **17** (2020), no. 1, 494–513.

[CFMW17] C. Conradi, E. Feliu, M. Mincheva, and C. Wiuf, *Identifying parameter regions for multistationarity*, PLoS computational biology **13** (2017), no. 10, e1005751.

[CFRS07] C. Conradi, D. Flockerzi, J. Raisch, and J. Stelling, *Subnetwork analysis reveals dynamic features of complex (bio)chemical networks*, Proceedings of the National Academy of Sciences **104** (2007), no. 49, 19175–19180.

[CHW08] G. Craciun, J.W. Helton, and R.J. Williams, *Homotopy methods for counting reaction network equilibria*, Mathematical biosciences **216** (2008), no. 2, 140–149.

[CL77a] M.D. Choi and T.Y. Lam, *Extremal positive semidefinite forms*, Mathematische Annalen **231** (1977), no. 1, 1–18.

[CL77b] ———, *An old question of Hilbert*, Queens papers in pure and applied mathematics **46** (1977), no. 385-405, 5.

[CLR95] M.D. Choi, T.Y. Lam, and B. Reznick, *Sums of squares of real polynomials*, Proceedings of Symposia in Pure Mathematics **58** (1995), 103–126.

[CLR02] ———, *Lattice polytopes with distinct pair-sums*, Discrete & Computational Geometry **27** (2002), no. 1, 65–72.

[CM14] C. Conradi and M. Mincheva, *Catalytic constants enable the emergence of bistability in dual phosphorylation*, Journal of The Royal Society Interface **11** (2014), no. 95, 20140158.

[CMS19] C. Conradi, M. Mincheva, and A. Shiu, *Emergence of oscillations in a mixed-mechanism phosphorylation system*, Bulletin of mathematical biology **81** (2019), no. 6, 1829–1852.

[Coh89] P. Cohen, *The structure and regulation of protein phosphatases*, Annual review of biochemistry **58** (1989), no. 1, 453–508.

[CS16] V. Chandrasekaran and P. Shah, *Relative entropy relaxations for signomial optimization*, SIAM Journal on Optimization **26** (2016), no. 2, 1147–1173.

[CS18] C. Conradi and A. Shiu, *Dynamics of post-translational modification systems: Recent progress and future directions*, Biophysical journal **114** (2018), no. 3, 507–515.

[DBMP14] P. Donnell, M. Banaji, A. Marginean, and C. Pantea, *CoNtRol: an open source framework for the analysis of chemical reaction networks*, Bioinformatics **30** (2014), no. 11, 1633–1634.

[DHNdW20] M. Dressler, J. Heuer, H. Naumann, and T. de Wolff, *Global optimization via the dual sonc cone and linear programming*, Preprint; see arXiv:2002.09368 (2020).

[DIdW17] M. Dressler, S. Iliman, and T. de Wolff, *A Positivstellensatz for sums of nonnegative circuit polynomials*, SIAM Journal on Applied Algebra and Geometry **1** (2017), no. 1, 536–555.

[DIdW19] ———, *An approach to constrained polynomial optimization via nonnegative circuit polynomials and geometric programming*, Journal of Symbolic Computation **91** (2019), 149–172.

[Die85] J.A. Dieudonne, *History of algebraic geometry*, Wadsworth Mathematics Series, Springer, 1985.

[DKdW18] M. Dressler, A. Kurpisz, and T. de Wolff, *Optimization over the boolean hypercube via sums of nonnegative circuit polynomials*, 43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018), vol. 117, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2018, p. 82.

[DNT18] M. Dressler, H. Naumann, and T. Theobald, *The dual cone of sums of nonnegative circuit polynomials*, Preprint; see arXiv:1809.07648 (2018).

[EFJK12] P. Ellison, M. Feinberg, H. Ji, and D. Knight, *Chemical reaction network toolbox, version 2.2*, Available online at `http://www.crnt.osu.edu/CRNTWin`, 2012.

[EKW00] M. El Kahoui and A. Weber, *Deciding hopf bifurcations by quantifier elimination in a software-component architecture*, J. Symbolic Computation **1** (2000), 1–19.

[FdW19] J. Forsgård and T. de Wolff, *The algebraic boundary of the sonc cone*, Preprint; see arXiv:1905.04776 (2019).

[Fei95] M. Feinberg, *The existence and uniqueness of steady states for a class of chemical reaction networks*, Archive for Rational Mechanics and Analysis **132** (1995), no. 4, 311–370.

[Fei19] _____, *Foundations of chemical reaction network theory*, Springer International Publishing, 2019.

[Fel15] E. Feliu, *Injectivity, multiple zeros and multistationarity in reaction networks*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **471** (2015), no. 2173.

[FHC14] D. Flockerzi, . Holstein, and C. Conradi, *n-site phosphorylation systems with 2n-1 steady states*, Bulletin of Mathematical Biology **76** (2014), no. 8, 1892–1916.

[FK10] C. Fidalgo and A. Kovacec, *Positive semidefinite diagonal minus tail forms are sums of squares*, Mathematische Zeitschrift **269** (2010), no. 3-4, 629–645.

[FKdWY20] E. Feliu, N. Kaihnsa, T. de Wolff, and O. Yürük, *The kinetic space of multistationarity in dual phosphorylation*, Journal of Dynamics and Differential Equations (2020).

[FSW+16] S. Feng, M. Sáez, C. Wiuf, E. Feliu, and O.S. Soyer, *Core signalling motif displaying multistability through multi-state enzymes*, Journal of The Royal Society Interface **13** (2016), no. 123, 20160524.

[FW12] E. Feliu and C. Wiuf, *Enzyme-sharing as a cause of multi-stationarity in signalling systems*, Journal of The Royal Society Interface **9** (2012), no. 71, 1224–1232.

[FW13] ———, *Variable elimination in post-translational modification reaction networks with mass-action kinetics*, Journal of mathematical biology **66** (2013), no. 1-2, 281–310.

[GH86] J. Guckenheimer and P. Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Applied Mathematical Sciences 42, Springer-Verlag New York, 1986.

[GJ00] E. Gawrilow and M. Joswig, `polymake`*: a framework for analyzing convex polytopes*, Polytopes—combinatorics and computation (Oberwolfach, 1997), DMV Sem., vol. 29, Birkhäuser, Basel, 2000, pp. 43–73. MR 1785292

[GM12] M. Ghasemi and M. Marshall, *Lower bounds for polynomials using geometric programming*, SIAM Journal on Optimization **22** (2012), no. 2, 460–473.

[GO04] J.E. Goodman and J. O'Rourke, *Handbook of discrete and computational geometry*, Discrete mathematics and its applications, Chapman & Hall/CRC, 2004.

[Gun03] J. Gunawardena, *Chemical reaction network theory for in-silico biologists*, Available online at `http://vcp.med.harvard.edu/papers/crnt.pdf`, 2003.

[GW95] M.X. Goemans and D.P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, Journal of the ACM **42** (1995), no. 6, 11151145.

[HF96] C.Y. Huang and J.E. Ferrell, *Ultrasensitivity in the mitogen-activated protein kinase cascade*, Proceedings of the National Academy of Sciences **93** (1996), no. 19, 10078–10083.

[Hil88] D. Hilbert, *Über die darstellung definiter formen als summe von formenquadraten*, Mathematische Annalen **32** (1888), no. 3, 342–350.

[Hil00] ———, *Mathematical problems*, Bull. Amer. Math. Soc. (N.S.) **37** (2000), no. 4, 407–436, Reprinted from Bull. Amer. Math. Soc. **8** (1902), 437–479.

[HLP34] G. H. Hardy, J.E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, 1934.

[HLS96] H. Hong, R. Liska, and S. Steinberg, *Testing stability by quantifier elimination*, Journal of Symbolic Computation (1996).

[HR15] J. Hell and A.D. Rendall, *A proof of bistability for the dual futile cycle*, Nonlinear Analysis: Real World Applications **24** (2015), 175–189.

[HR17] ———, *Dynamical features of the MAP kinase cascade*, Modeling Cellular Systems, Springer, 2017, pp. 119–140.

[HRdWY20] J. Hartzer, O. Röhrig, T. de Wolff, and O. Yürük, *Initial steps in the classification of maximal mediated sets*, Journal of Symbolic Computation (2020).

[HT81] V. Hárs and J. Tóth, *On the inverse problem of reaction kinetics*, Qualitative Theory of Differential Equations **30** (1981), 363–379.

[Hur91] A. Hurwitz, *Ueber den Vergleich des arithmetischen und des geometrischen Mittels.*, Journal für die reine und angewandte Mathematik **108** (1891), 266–268.

[IdW16a] S. Iliman and T. de Wolff, *Amoebas, nonnegative polynomials and sums of squares supported on circuits*, Research in the Mathematical Sciences **3** (2016), no. 9, 1–35.

[IdW16b] ———, *Lower bounds for polynomials with simplex newton polytopes based on geometric programming*, SIAM Journal on Optimization **26** (2016), no. 2, 1128–1146.

[Kar72] R.M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations, Springer, 1972, pp. 85–103.

[KdW19] A. Kurpisz and T. de Wolff, *New dependencies of hierarchies in polynomial optimization*, Proceedings of the 2019 on International Symposium on Symbolic and Algebraic Computation, 2019, pp. 251–258.

[KFCS15] V.B. Kothamachu, E. Feliu, L. Cardelli, and O.S. Soyer, *Unlimited multistability and boolean logic in microbial signalling*, Journal of the Royal Society interface **12** (2015), no. 108.

[KNT19] L. Katthän, H. Naumann, and T. Theobald, *A unified framework of sage and sonc polynomials and its duality theory*, Preprint; see arXiv:1903.08966 (2019).

[Kri64] J.-L. Krivine, *Anneaux préordonnés*, Journal danalyse mathématique **12** (1964), no. 1, 307–326.

[Lai78] K.J. Laidler, *Physical chemistry with biological applications*, Benjamin/Cummings Pub. Co., 1978.

[Lan] *Algebra*, Graduate texts in Mathematics.

[Las01] J.B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on optimization **11** (2001), no. 3, 796–817.

[Las10] ———, *Moments, positive polynomials and their applications*, vol. 1, World Scientific, 2010.

[Lau09] M. Laurent, *Sums of squares, moment matrices and optimization over polynomials*, Emerging applications of algebraic geometry, Springer, 2009, pp. 157–270.

[LK99] M. Laurent and N. Kellershohn, *Multistability: a major means of differentiation and evolution in biological systems*, Trends in biochemical sciences **24** (1999), no. 11, 418–422.

[Mar08] M. Marshall, *Positive polynomials and sums of squares*, no. 146, American Mathematical Society, 2008.

[MCW20a] R. Murray, V. Chandrasekaran, and A. Wierman, *Newton polytopes and relative entropy optimization*, Preprint; see arXiv:1810.01614 (2020).

[MCW20b] ———, *Signomial and polynomial optimization via relative entropy and partial dualization*, Mathematical Programming Computation (2020), 1–39.

[MDSC12] M.P. Millán, A. Dickenstein, A. Shiu, and C. Conradi, *Chemical reaction systems with toric steady states*, Bulletin of Mathematical Biology **74** (2012), no. 5, 1027–1065.

[Mes82] B.E. Meserve, *Fundamental concepts of algebra*, Dover Publications, 1982.

[MHK04] N.I. Markevich, J.B. Hoek, and B.N. Kholodenko, *Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades*, The Journal of cell biology **164** (2004), no. 3, 353–359.

[MK87] K.G. Murty and S. Kabadi, *Some np-complete problems in quadratic and nonlinear programming*, Mathematical Programming **39** (1987), no. 2, 117–129.

[Mot67] T.S. Motzkin, *The arithmetic-geometric inequality*, Inequalities: Proceedings, Volume 1, Academic Press, 1967, pp. 203–224.

[NN94] Y. Nesterov and A.S. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, Society for Industrial and Applied Mathematics, 1994.

[OBS99] M.A. Olson, K. Bostic, and M. Seltzer, *Berkeley DB*, Proceedings of the Annual Conference on USENIX Annual Technical Conference (Berkeley, CA, USA), ATEC '99, USENIX Association, 1999, pp. 43–43.

[OTL+04] E.M. Ozbudak, M. Thattai, H.N. Lim, B.I. Shraiman, and A. Van Oudenaarden, *Multistability in the lactose utilization network of escherichia coli*, Nature **427** (2004), no. 6976, 737–740.

[Oxl11] J.G. Oxley, *Matroid theory*, Oxford Graduate Texts in Mathematics, 21, Oxford University Press, 2011.

[Par00] P.A. Parrilo, *Structured semidefinite programsand semialgebraic geometry methods in robustness and optimization*, Ph.D. thesis, California Institute of Technology, 2000.

[Par03] ———, *Semidefinite programming relaxations for semialgebraic problems*, Mathematical Programming **96** (2003), no. 2, 293320.

[PD13] A. Prestel and C. Delzell, *Positive polynomials: from hilberts 17th problem to real algebra*, Springer Science & Business Media, 2013.

[PMD18] M. Pérez Millán and A Dickenstein, *The structure of MESSI biological systems*, SIAM Journal on Applied Dynamical Systems **17** (2018), no. 2, 1650–1682.

[PR20] V. Powers and B. Reznick, *A note on mediated simplices*, 2020, p. 106608.

[Put93] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana University Mathematics Journal **42** (1993), no. 3, 969–984.

[PW98] V. Powers and T. Wörmann, *An algorithm for sums of squares of real polynomials*, Journal of Pure and Applied Algebra **127** (1998), no. 1, 99–104.

[QNKS07] L. Qiao, R.B. Nachbar, I.G. Kevrekidis, and S.Y. Shvartsman, *Bistability and oscillations in the Huang-Ferrell model of MAPK signaling*, PLoS Computational Biology **3** (2007), no. 9, 1819–1826.

[Rez78] B. Reznick, *Extremal psd forms with few terms*, Duke Mathematical Journal **45** (1978), no. 2, 363–374.

[Rez89] ———, *Forms derived from the arithmetic-geometric inequality*, Mathematische Annalen **283** (1989), no. 3, 431–464.

[Rez00] ———, *Some concrete aspects of hilbert's 17th problem*, Contemporary Mathematics **253** (2000), 251–272.

[Rob69] R.M. Robinson, *Some definite polynomials which are not sums of squares of real polynomials*, Notices of the American Mathematical Society **16** (1969), no. 3, 554.

[Roe20] O. Roehrig, *Initial steps in the classification of maximal mediated sets*, Master's thesis, TU Berlin, 2020.

[Sch91] K. Schmüdgen, *The k-moment problem for compact semi-algebraic sets*, Mathematische Annalen **289** (1991), no. 1, 203–206.

[Sch11] A. Schrijver, *Theory of linear and integer programming*, Wiley, 2011.

[SdW18] H. Seidler and T. de Wolff, *An experimental comparison of SONC and SOS certificates for unconstrained optimization*, Preprint; see arXiv:1808.08431 (2018).

[SdW19] ———, *POEM: Effective methods in polynomial optimization, version 0.2.1.0(a)*, `http://www.iaa.tu-bs.de/AppliedAlgebra/POEM/index.html`, jul 2019.

[Ste74] G. Stengle, *A nullstellensatz and a positivstellensatz in semialgebraic geometry*, Mathematische Annalen **207** (1974), no. 2, 87–97.

[Ste10] J.M. Steele, *The cauchy-schwarz master class: an introduction to the art of mathematical inequalities*, Cambridge University Press, 2010.

[Tan11] O. Tange, *GNU parallel - the command-line power tool*, ;login: The USENIX Magazine **36** (2011), no. 1, 42–47.

[TF20] A. Torres and F. Feliu, *Symbolic proof of bistability in reaction networks*, 2020.

[TG09a] M. Thomson and J. Gunawardena, *The rational parameterisation theorem for multisite post-translational modification systems*, Journal of Theoretical Biology **261** (2009), no. 4, 626–636.

[TG09b] ———, *Unlimited multistability in multisite phosphorylation systems*, Nature **460** (2009), no. 7252, 274–277.

[VB96] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Review **38** (1996), no. 1, 49–95.

[W.20] Jie W., *Nonnegative polynomials and circuit polynomials*, Preprint; see arXiv:1804.09455 (2020).

[Wan04] X. Wang, *A simple proof of descartes's rule of signs*, The American Mathematical Monthly **111** (2004), no. 6, 525.

[WF13] C. Wiuf and E. Feliu, *Power-law kinetics and determinant criteria for the preclusion of multistationarity in networks of interacting species*, SIAM Journal on Applied Dynamical Systems **12** (2013), no. 4, 1685–1721.

[WS08] L. Wang and E.D. Sontag, *On the number of steady states in a multiple futile cycle*, Journal of mathematical biology **57** (2008), no. 1, 29–52.

[WSV03] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of semidefinite programming: theory, algorithms and applications*, Kluwer Academic, 2003.

[XF03] W. Xiong and J.E. Ferrell, *A positive-feedback-based bistable 'memory module' that governs a cell fate decision*, Nature **426** (2003), no. 6965, 460–465.

[Zie95] G.M. Ziegler, *Lectures on polytopes*, Graduate texts in mathematics, 152, Springer-Verlag, 1995.